


The Neural Basis for Combinatorial Coding in a Cortical Population Response

Leslie C. Osborne,^{1*} Stephanie E. Palmer,^{3*} Stephen G. Lisberger,^{1,2} and William Bialek³ 

¹Sloan-Swartz Center for Theoretical Neurobiology, W. M. Keck Foundation Center for Integrative Neuroscience, and Department of Physiology, and

²Howard Hughes Medical Institute, University of California at San Francisco, San Francisco, California 94143, and ³Joseph Henry Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544

We have used a combination of theory and experiment to assess how information is represented in a realistic cortical population response, examining how motion direction and timing is encoded in groups of neurons in cortical area MT. Combining data from several single-unit experiments, we constructed model population responses in small time windows and represented the response in each window as a binary vector of 1s or 0s signifying spikes or no spikes from each cell. **We found that patterns of spikes and silence across a population of nominally redundant neurons can carry up to twice as much information about visual motion than does population spike count,** even when the neurons respond independently to their sensory inputs. This extra information arises by virtue of the **broad diversity of firing rate dynamics found in even very similarly tuned groups of MT neurons.** Additionally, **specific patterns of spiking and silence can carry more information than the sum of their parts (synergy), opening up the possibility for combinatorial coding in cortex.** These results also held for populations in which we imposed levels of nonindependence (correlation) comparable to those found in cortical recordings. Our findings suggest that combinatorial codes are advantageous for representing stimulus information on **short time scales,** even when neurons have no complicated, stimulus-dependent correlation structure. 

Key words: visual motion; MT; spike trains; information theory; smooth pursuit; correlated variability

Introduction

The modern study of neural coding began with the discovery that neural responses to sensory stimuli are composed of discrete events: either a cell generates an action potential or it remains silent (Adrian, 1926). If we average, either over long periods of time or over many presentations of the same sensory inputs, then we can recover a continuous description of neural responses in terms of the firing rate or probability of spiking. **But the brain does not have access to averages across long times or repeated presentations;** behaviors occur in response to individual stimuli and often are evoked by changes in sensory input that occur on very short time scales. The mammalian brain can, however, **average responses over a population of neurons** and has access to many cortical neurons that are nominally redundant in the sense that they have nearly identical feature selectivity (Mountcastle, 1957, 1997; Hubel and Wiesel, 1962). In this study, we ask whether averaging is the best strategy for pooling information from a nominally redundant population, or whether the under-

lying discreteness of spikes and silences is informative even in large populations of neurons.

In general, populations of neurons with similar feature selectivity could function in two very different ways. In one view, the signals carried by the different cells are redundant and primarily independent, so that averaging over the population serves to reduce noise. At the opposite extreme, information is transmitted by a combinatorial code, defined as the pattern of spikes and silences across the population. **If different combinations of spikes and silences stand for different sensory stimuli, then averaging would discard potentially large amounts of information present only in the combinatorial code.** A number of groups have presented evidence for combinatorial coding, in the cortex (Gawne and Richmond, 1993; Panzeri et al., 1999; Reich et al., 2001; Narayanan et al., 2005; Montani et al., 2007) and elsewhere (Purcell and Bullock, 1968; Schneidman et al., 2006b), but it seems fair to say that the view of averaging for noise reduction continues to be more prominent in discussions of sensory coding in the cortex.

Here we reexamine the issue of combinatorial coding using **neural responses to visual motion recorded in the middle temporal (MT) area of visual cortex** (Dubner and Zeki, 1971; Maunsell and van Essen, 1983; Albright et al., 1984). In MT, as in other cortical areas, it is difficult to record simultaneously from more than a few neurons with similar feature selectivity. We have extended previous work by using sequential recordings from many MT cells to construct models that are faithful to the measured properties of the individual cells but allow us to extrapolate to the behavior of larger populations. Even when we assume that cells

Received Sept. 14, 2008; revised Oct. 23, 2008; accepted Oct. 27, 2008.

This work was supported by National Institutes of Health Grants EY017210 and EY03878, the Howard Hughes Medical Institute, the Life Sciences Research Foundation, and the Swartz Foundation. We thank Stefanie Tokiyama and Karen McLeod for assistance with animal monitoring and maintenance, Scott Ruffner for software development, and Nicholas Priebe and Carlos Cassanello for participating in the original experiments.

*L.C.O. and S.E.P. contributed equally to this work.

Correspondence should be addressed to Stephanie E. Palmer, Department of Physics, Princeton University, Lewis-Sigler Institute for Integrative Genomics, Carl Icahn Laboratory, Room 238, Princeton, NJ 08544. E-mail: sepalmer@princeton.edu.

DOI:10.1523/JNEUROSCI.4390-08.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/2813522-10\$15.00/0


respond independently to their sensory inputs, the situation when averaging works best, we find that particular combinatorial patterns of spiking and silence in such populations frequently carry substantial information about the sensory input, information that is lost by averaging across neurons. This coding advantage is robust over a reasonable range of correlation strengths and arises because neurons can have very different response dynamics, even when they are nominally redundant in the sense that they possess similar feature selectivity. **Our central conclusion is that combinatorial coding becomes possible as soon as cells having “similar feature selectivity” along conventional axes (such as motion direction and speed in MT) possess some diversity of response properties along another axis.**

Materials and Methods

Experimental methods. Experimental data have been published previously (Osborne et al., 2004). To acquire these data, extracellular single-unit microelectrode recordings were made in three **sufentanil-anesthetized, paralyzed monkeys (*Macaca fascicularis*)** according to a protocol that had been approved in advance by the Institutional Animal Care and Use Committee at the University of California at San Francisco. Using **random dot texture stimuli** presented on a high-resolution analog oscilloscope display, we mapped receptive field location, determined the preferred direction and speed of the neuron under study, and sized the stimuli to maximally excite each neuron. The random dot texture was moved behind a stationary aperture, creating a moving stimulus at a fixed retinal location.

Visual stimuli were presented in discrete trials. **Each stimulus appeared and remained stationary for 256 ms, then stepped to a constant velocity for 256 ms, and was again stationary for 256 ms.** A brief pause separated successive trials, and directions of motion were pseudorandomly interleaved. A typical experiment included 13 motion directions that spanned $\pm 90^\circ$ around the neuron's preferred direction in 15° increments, all presented at the neuron's preferred speed. Each stimulus was presented up to 222 times. Spike times were recorded with 10 μ s resolution.

Constructing a model population. We consider first a model population in which each cell responds independently to its sensory inputs. Mathematically, this means that the probability of responses from the population can be decomposed as a product of probabilities for each individual cell, as in Equation 1 below, where these single-cell properties have been estimated directly from the experimental data. Examples of the responses from the population can be drawn out of this distribution, but for small groups of cells, we can also draw samples directly from the raw data. These approaches are mathematically equivalent.

To create a model population of 10 or fewer neurons  nominally redundant feature selectivity, we aligned all cells by their preferred direction and **assumed that all neurons had the same preferred speed.** Then, we resampled the rasters of individual cells at $\Delta\tau = 8$ ms resolution, labeling the occupancy of each time bin with a “1” if there had been one or more spikes in the time interval or a “0” if there had not. At 8 ms resolution, very few neurons emitted multiple spikes in a single bin; this occurred in fewer than 10% of the bins in our analyses. Thus, 8 ms was a good compromise between filling bins with spikes but not overfilling them with multiple spike events. Rather than focus on the pattern of spikes and silences across time in individual neurons, **we created binary population “words,” patterns of 1s and 0s across neurons in a group, and used the words to represent whether each neuron had fired or not in a single time bin.** Hence, at each time point during the response, we drew the N letters of each word randomly from the collection of stimulus repetitions from all N cells in our sample in the appropriate bin. Each neuron in our sample corresponded to a fixed position in the word, and we could construct many different words by random draws from the many repetitions of each stimulus for each neuron. The probability of observing a particular word, $P(\mathbf{n} \equiv \{n_i\})$, where n_i labels the count (0 or 1) in neuron i , was then measured by estimating the frequency of occurrence of that pattern of 1s and 0s within the entire dataset.

Directly sampling $P(\mathbf{n})$ from the independently recorded neural responses in this way is tractable for small populations of fewer than 10 neurons. For population sizes >10 , however, the intuitive method of resampling from the data requires generation of millions of samples to avoid introducing biases in the estimation of information theoretic quantities. Therefore, rather than sampling the rasters directly for groups larger than 10, we construct the probability distributions for the responses using the trial-averaged firing rates of the individual neurons.

On the assumption that our neurons respond independently to their sensory inputs, we can write the conditional probability distribution for the population's response $\mathbf{n} \equiv n_i$ at some moment of time t (defined relative to stimulus onset) as follows:

$$P(\mathbf{n}|t) = \prod_{i=1}^N P(n_i|t). \quad (1)$$

In terms of the spiking probabilities of each neuron, in the case in which either one spike ($n_i = 1$) or no spikes ($n_i = 0$) are emitted in each time bin, this distribution takes the following form:

$$P(\mathbf{n}|t) = \prod_{i=1}^N [q_i(t)]^{n_i} [1 - q_i(t)]^{1-n_i}, \quad (2)$$

where $q_i(t)$ denotes the probability of a spike from the i th neuron at time t . The expression can be rewritten as follows:

$$P(\mathbf{n}|t) = \frac{1}{Z(t)} \exp \left[\sum_{i=1}^N \phi_i(t) n_i \right], \quad (3)$$

where $Z(t)$ is a normalization constant and $\phi_i(t) = \ln \left(\frac{q_i(t)}{1 - q_i(t)} \right)$. In the limit $\Delta\tau \rightarrow 0$, we can identify the time-dependent firing rate of each neuron $r_i(t) = q_i(t)/\Delta\tau$. Thus, the instantaneous firing rates of our cells fully determine the $\phi_i(t)$ s, and we can compute the word probabilities, $P(\mathbf{n})$, using Equation 3 without any further assumptions. This computational approach works because the spike rate of each neuron, at each moment of time, is well determined by our experimental data with small error bars. Using this method allows us to estimate the word distribution for large population sizes in which direct sampling would be difficult both computationally and experimentally. We checked that the method yielded the same answers as the sampling approach for values of N where the calculations were tractable.

Constructing a correlated population. We continue to work in small time windows, of duration $\Delta\tau = 8$ ms, such that the response of each neuron i consists either of one spike ($n_i = 1$) or silence ($n_i = 0$). **In the independent model (Eq. 3), there are no correlations between the responses of different neurons, n_i and n_j , once we know the stimulus.** We are able to construct correlations among neurons by adding an explicit term to the exponential that couples the responses of the different cells:

$$P(\mathbf{n}|t) = \frac{1}{Z(t)} \exp \left[\sum_{i=1}^N \phi_i(t) n_i + \frac{J}{2} \sum_{i=1}^N \sum_{j \neq i}^N n_i n_j \right]. \quad (4)$$

Equation 4 defines a family of models with a range of correlations between our neurons while preserving their experimentally observed time-varying firing rates. For small J , the strength of these correlations is proportional to J . This coupling is constant in time and stimulus independent. **Equation 4 is the least structured or maximum entropy model** that generates some average level of correlations among all the pairs of neurons (Jaynes, 1957; Schneidman et al., 2003; Schneidman et al., 2006a).

If we choose some value of the coupling, J , which ultimately will set the strength of correlations in the population, we still need to be sure that, at every moment, the probability of spiking from the individual cells matches the observed firing rates as a function of time $r_i(t)$. To do this, we solve for the $\phi_i(t)$ s in Equation 4, subject to the following constraints:

$$\sum_n n_k(t) P(\mathbf{n}|t) = r_k(t) \Delta\tau, \quad (5)$$

where the $r_k(t)$ s are measured single-cell firing rates, averaged over a small time window, $\Delta\tau = 8$ ms. Because these equations are not coupled in time, we can solve for the fields at each time point independently. We have an analytical solution for the fields with $J = 0$, and we can proceed from this solution using perturbation theory, from which we obtain an equation relating small changes in J to their effect on the fields, ϕ :

$$\Delta\phi_i = \left(\frac{\Delta J}{2} \sum_{\alpha \neq \beta} \langle n_\alpha n_\beta \rangle \langle n_k \rangle - \langle n_\alpha n_\beta n_k \rangle \right) \chi_{ik}^{-1}, \quad (6)$$

where α and β index neurons in the group of N cells and χ is the connected part of the two-point correlation function, $\chi = \langle n_i n_k \rangle - \langle n_i \rangle \langle n_k \rangle$. In practice, we solved for the fields at very small increments, $\Delta J/J = 0.001$, checking satisfaction of the constraints on the firing rates at each step. This perturbative approach is fast but accumulates errors. To correct for the accumulated errors, we performed local function minimization whenever the fractional error in the single-cell rates exceeded 10^{-8} and then returned to the perturbative stepping until the error bound again was reached.

Once we created a model population response, we summed the spike counts across the full time window of the response to motion and computed the correlation coefficients between counts in all pairs of neurons in our model population. The mean of these coefficients provides an index for the overall strength of the correlations. Experimentally, for neurons in MT, the correlation coefficients are in the range of 0.1–0.2 (van Kan et al., 1985; Gawne and Richmond, 1993; Zohary et al., 1994; Lee et al., 1998; Bair et al., 2001; Averbeck and Lee, 2003; Kohn and Smith, 2005), which corresponds to $J = 0.11$ to 0.16 in our models.

Estimating information. To estimate the information carried by population words about the stimulus, we first computed the probability of observing each particular N -neuron word from our dataset $P(\mathbf{n} \equiv \{n_i\})$, where n_i labels the count (0 or 1) in neuron i . Word probabilities were determined by the frequency of occurrence across all motion directions and times. The **total entropy** of the words is given by the following:

$$S[P(\mathbf{n})] = - \sum_{\mathbf{n}} P(\mathbf{n}) \log_2 P(\mathbf{n}). \quad (7)$$

The probability of observing a word for a particular stimulus, $P(\mathbf{n}|\phi, t)$, was estimated in a similar manner to $P(\mathbf{n})$, but now at each particular time t relative to the onset of motion in a direction ϕ . The entropy of the conditional distributions is given by the following:

$$S[P(\mathbf{n}|\theta, t)] = - \sum_{\mathbf{n}} P(\mathbf{n}|\theta, t) \log_2 P(\mathbf{n}|\theta, t). \quad (8)$$

The **average amount of information** that words carry about the stimulus is given by the difference between the total entropy and the average noise entropy:

$$I_{\text{words}} = S[P(\mathbf{n})] - \left(\frac{1}{13 \cdot T} \sum_t \sum_{\theta} S[P(\mathbf{n}|\theta, t)] \right), \quad (9)$$

where T represents the total number of time bins in the response and ϕ indexes the **13 motion directions**. We computed the information from counts in a similar way, using the same $P(\mathbf{n})$ as before, but collapsing over words with the same number of spikes, such that $P_{\text{count}}(n) = \sum_{\mathbf{n}} P(\mathbf{n}) \cdot \delta(n, \sum_i n_i)$, where the sum runs over all words, \mathbf{n} , that have a count equal to n , enforced by the δ function, $\delta(n, \sum_i n_i)$, which is unity when the sum equals n and zero otherwise. Similarly, $P_{\text{count}}(n|\theta, t) = \sum_{\mathbf{n}} P(\mathbf{n}|\theta, t) \cdot \delta(n, \sum_i n_i)$. With $P_{\text{count}}(n)$ in hand, we computed I_{counts} in a completely analogous manner to the calculation of I_{words} .

When we computed word probabilities directly from Equation 3, we propagated the errors in measured firing rates to obtain errors in the derived information measures. In cases in which we generated samples of population words directly from observed spike trains, all entropy estimates were corrected for finite sampling effects by taking multiple random samples of fractions of the dataset and then performing a linear extrapolation to infinite sample size (Strong et al., 1998). Errors in infor-

mation quantities were estimated by extrapolating the SD of values computed from half the sample in the same manner.

All of our information quantities depend only on patterns of spiking and silence in single time bins. Thus, for the purpose of our calculations, correlations between time bins (and hence the question of whether the spike trains approximate modulated Poisson processes) are irrelevant. As a test of our computations, we created shuffled spike trains with exact Poisson statistics and reproduced all of our results.

Testing the effects of population response diversity: tuning versus dynamics. After analyzing the information available from different groups of real neural responses, we asked how the information from words changed when the responses of different populations of neurons were altered artificially to have the same direction tuning width and/or the same time-varying firing rate. To do so systematically, we repeated the analysis many times, using the direction tuning and time-varying firing rate of each neuron, in turn, as a template. For each group of 10 cells, we randomly chose another neuron from the population as the template and computed its normalized tuning curve, $\bar{r}(\theta)^* = \langle r(t, \theta) \rangle_t / \langle r(t, \theta = 0) \rangle_t$, where the asterisk indicates the template neuron, the bar indicates normalization, and $\langle \dots \rangle_t$ indicates a time average over the 256 ms presentation of one motion direction, θ . We also computed the shape of the selected neuron's temporal modulations in rate at the preferred direction, $\bar{r}(t)^* = r(t, \theta = 0) / \langle r(t, \theta = 0) \rangle_t$. We then used these two functions to serve as templates for fixing the tuning or firing rate dynamics of the group. To fix the tuning of the population, we allowed each cell to retain its own temporal dynamics, but rescaled each trace, denoted $\bar{r}(t, \theta)$, by a constant factor that forced the cell's tuning curve to have the same shape as $\bar{r}(\theta)^*$, such that $\bar{r}(t, \theta) = [\langle r(t, \theta) \rangle_t / \langle r(t, \theta = 0) \rangle_t] \cdot \bar{r}(\theta)^*$. Note that rescaling fixes the tuning curve shape (i.e., the bandwidth) but allows each neuron's tuning curve to retain its original peak amplitude. To fix the firing rate dynamics, each cell retained its own directional tuning curve, but temporal dynamics were set by the template $\bar{r}(t)^*$, so that $\bar{r}(t, \theta)^* = \langle r(t, \theta) \rangle_t \cdot \bar{r}(t)^*$.

Spike-silence synergy. To measure the synergy between spikes and silences in our population words, we took the difference between the stimulus information that a particular word captured and the sum of the information from each component spike and silence (Brenner et al., 2000; Schneidman et al., 2006b),

$$I_{\text{synergy}} = I(\{n_i\}) - \sum_i I(n_i). \quad (10)$$

The stimulus information, I , was computed as described by Brenner et al. (2000) and is given by the following:

$$I(\text{stimulus}) = \frac{1}{T} \int_0^T dt \frac{r(t)}{\langle r(t) \rangle_t} \log_2 \left(\frac{r(t)}{\langle r(t) \rangle_t} \right), \quad (11)$$

where $r(t)$ is the rate for a given event, the occurrence of a given word, or a spike or silence from a particular cell.

Plotting response–conditional stimulus distributions. We measured the distribution of stimuli giving rise to a particular response pattern directly from our neural data. We then smoothed the result with a Gaussian kernel, the parameters of which were chosen to approximate the size of a “pixel” determined by the direction spacing used in our experiments and the chosen temporal resolution. In this kernel approach, we approximate the distribution as a (normalized) sum of terms, in which each term is a small “blob” centered on each stimulus that gives rise to the particular neural response pattern. Here we are estimating a distribution of motion direction and time since motion onset, and we used a two-dimensional Gaussian kernel, so that our estimate becomes as follows:

$$P(\theta, t|\mathbf{n}) \propto \sum_{\mu} \exp(-[(t - t_{\mu})^2 / (2\sigma_t^2) + (\theta - \theta_{\mu})^2 / (2\sigma_{\theta}^2)]), \quad (12)$$

where $\sigma_t = 20$ ms and $\sigma_{\theta} = 10^\circ$. The values of σ_t and σ_{θ} were chosen to be about the size of one time or direction bin so that we obtained a

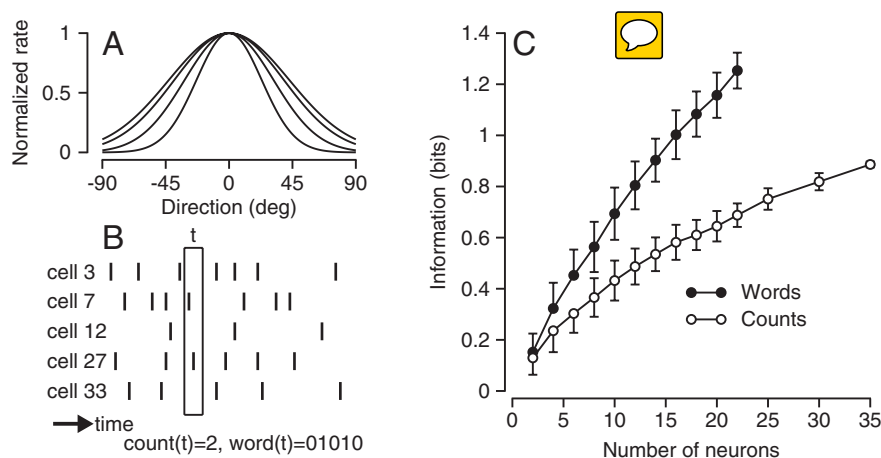


Figure 1. Utility of patterns of spiking and silence across a diverse population of MT neurons for providing information about a target motion stimulus. **A**, The normalized tuning curves of four MT neurons, showing firing rate versus direction of motion. Data were normalized by the response to the preferred direction, relative to which all other directions are measured. deg, Degrees. **B**, Method for creating words to indicate spiking and silences across a population of neurons, in 8 ms windows. In this example, the population response at time t is characterized by the word “01010.” **C**, The information that counts and words carry about the visual motion stimulus plotted as a function of the number of neurons in the analysis population. Error bars indicate SD.

smoothed estimate of the conditional distribution but did not lose resolution.

Results

Given a population of neurons that, by conventional criteria, have the same feature selectivity, we want to know whether averaging over the population makes the best use of the neural responses, or whether averaging discards information that is embedded in a combinatorial code. Our data comprise recordings from extrastriate cortical area MT, which provides the sensory inputs that mediate both the perception of visual motion and behaviors like smooth pursuit eye movements (Newsome et al., 1985; Groh et al., 1997; Born et al., 2000). Guided by the time scales characterizing the dynamics and precision of smooth pursuit (Osborne et al., 2005, 2007; Osborne and Lisberger, 2007), we focus on neural responses in small time windows, during which time each cell typically generates either a single action potential (1) or not (0); patterns of response across the population thus consist of binary words, whereas the averaged or pooled response of the population is given simply by the total number of spikes. There are many plausible but distinct definitions of “combinatorial coding.” In this section of the paper, we leave the abstract definition of combinatorial coding aside and focus on a series of concrete questions about the relationship of words and counts to the sensory input: Do the words carry information about the stimulus that is lost when we keep only the spike counts? Why? Do different words with the same number of spikes stand for different sensory inputs? Is the information provided by the pattern of spikes and silences greater than the sum of the contributions of each spike or silence in the individual cells? In the Discussion, we argue that the answers to these questions give us a new perspective on the problem of combinatorial coding in cortical area MT and in neural populations more generally.

Extra information about stimulus properties from patterns of spikes

To evaluate the possible existence of a combinatorial code, we consider a population of N MT neurons. Experimentally, we have observed many responses to each of a finite set of different stimuli (Osborne et al., 2004). Each of the cells in our sample was direc-

tionally tuned, with relatively similar selectivity and bandwidth when responses were normalized and preferred directions were aligned (Fig. 1A); by conventional measures, our sample is a redundant population of neurons. If we assume that each neuron responds independently to its sensory inputs, we can draw a single trial response from each neuron in our database to create a model population response, although the samples were recorded sequentially from many different neurons (see Materials and Methods for details). Armed with the model population responses, we can answer the questions outlined above.

If we look in a small window of time $\Delta\tau$, then the i th cell generates n_i spikes, with $i = 1, 2, \dots, N$. For small values of $\Delta\tau$, we will almost never see two spikes from a single cell. Thus, the response of the population $\{n_i\}$ at that moment in time can be treated as an N -letter binary word, \mathbf{n} (representing a “vector” pattern of 1s and 0s across labeled cells), as shown in Figure

1B. By keeping track of the appearance of combinations of spiking and silence across the population over all times in the response, we can ask how much information these code words carry about the stimulus. At each instant of time, the stimulus in our experiments is specified by the direction of motion θ and the time from the onset of motion, $t - t_{\text{onset}}$, and calculations described in Materials and Methods, allow us to use the experimental data to estimate the information that the pattern or number of spikes provides about the stimulus, $I(\mathbf{n}; \theta, t - t_{\text{onset}})$ or $I(n; \theta, t - t_{\text{onset}})$, respectively.

The results in Figure 1C demonstrate that the information provided by binary code words \mathbf{n} increases as a function of the number of cells N that contribute to the word, exceeding 1 bit for a population of 16 neurons. If we use the same draws from the experimental data to estimate the information that the (scalar) spike count $n \equiv \sum_{i=1}^N n_i$ provides about the stimulus, we find that the total amount of information from spike counts is smaller than the total information from words and never exceeds 1 bit even when spike counts are pooled across all neurons. The combinations of spiking and silence in the model MT population provide more than twice as much information as the pooled spike counts, although the cells in the pool have nominally redundant feature selectivity. Our first finding extends the previous analysis of Reich et al. (2001) from 6 neurons to 22 neurons, showing that a combinatorial code could be advantageous even for large neural populations. In this analysis, it is important that the stimulus is defined both by the direction of motion and by the time since motion onset; even with a limited number of motion directions, the availability of information about timing avoids the possibility that information will saturate because the set of inputs is too impoverished, as discussed by Rieke et al. (1997) and by Reich et al. (2001).

The fact that patterns of activity provide extra information about the stimulus compared with counting spikes means that the neurons are not fully redundant. Although they have very similarly shaped tuning curves and we have treated them as preferring the same direction of stimulus motion, they vary in response amplitude, in tuning bandwidth, and in the time course of the trial-averaged firing rate during the response to a step of

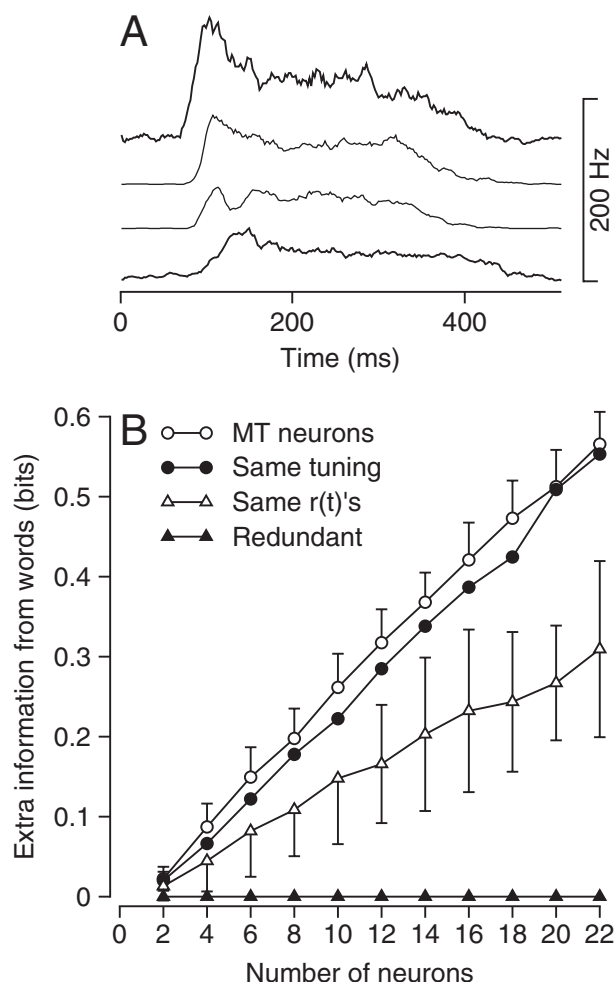


Figure 2. Combinatorial coding is enabled by a diversity of response dynamics. **A**, Responses of the same four neurons as in Figure 1A, plotting rate (from peristimulus time histogram) versus time in response to a 256 ms step of target speed in the preferred direction. Firing rate curves have been offset horizontally to improve visibility. **B**, Extra information from words versus counts is plotted as a function of the number of neurons in the analysis population. Different symbols show the results from populations of real and model neurons with different features made redundant. Open circles, Data drawn from actual single-trial responses; filled circles, diversity of response dynamics in model neurons mimics that in the data, but each neuron has been made to have the same tuning curve width; open triangles, model neurons that have the same time-varying firing rate, but response amplitude varies as in the actual data; filled triangles, model neurons that have the same time-varying firing rate and direction-dependent response amplitudes but are independent Poisson processes. Error bars indicate SD.

target velocity (Lisberger and Movshon, 1999). As illustrated in Figure 2A, the firing rate can rise quickly or slowly, and different neurons might show a small transient, a larger transient, or a purely tonic response. To ascertain which feature of the neural response provides the extra stimulus information available from words versus pooled spike counts, we next created a number of carefully contrived populations of 10 model neurons that preserved either the diversity of time-varying firing rates or the diversity of direction tuning curves, or that eliminated all diversity (see Materials and Methods). For each population, we then performed the same set of information calculations that led to Figure 1C.

For a population of model units that preserved the diversity of firing rate dynamics, $r(t)$, but forced all the neurons to have the same direction tuning width (Fig. 2B, filled circles), the amount of extra information from words was the same as that for the draws from the experimentally observed spike trains of MT neu-

rons (open circles). If we contrived each unit to have the same time-varying trajectory of firing rate $r(t)$, but retained the diversity of directional tuning bandwidths, then about half of the extra information from words was lost (Fig. 2B, open triangles). Thus, the diversity of firing rate dynamics, and not diversity of tuning curve bandwidth, accounts for much of the information advantage of words over counts. The extra information that remains reflects the fact that differences in the response amplitude at the peak of the tuning curve across neurons (Maunsell and van Essen, 1983) impose different time-averaged absolute firing rates across the population, even if the shape of the trajectory of the trial-averaged firing rate $r(t)$ was the same for each model neuron. We note that similar results on this latter point were obtained by Shamir and Sompolsky (2006), examining the effects of simulated heterogeneities in static tuning on population codes. Finally, if we created populations of fully redundant model units with one uniform trajectory, $r(t)$, and the same direction tuning curves, then the extra information from words was lost, as expected (Fig. 2B, filled triangles). Analysis of information as a function of time revealed that the extra information in the combinatorial code was concentrated near the time of the onset transients of the neural response, where the diversity of response dynamics is greatest (data not shown).

What does the extra information tell us?

The results of the previous section tell us how much information the patterns of spiking and silence can convey about the stimulus. The next step is to understand what these patterns are telling us about the stimulus. To focus our attention on a manageable set of patterns, we first computed the information carried by words and counts for different total spike counts in populations of $N = 2 \dots 16$ cells, drawing 100 populations for each condition from our 36 cortical neurons. Figure 3A shows that, for these population sizes, most of the extra information carried by words versus counts comes from words with relatively few spikes (i.e., from analysis windows when only a few of the neurons in the population emitted spikes and the rest were silent). Furthermore, most words had zero, one, or two spikes, and increasing numbers of spikes were progressively less common (Fig. 3B). Combining these two effects shows that the dominant term in the extra information provided by words typically comes from instances when only one neuron fired a spike. Even when the size of the population was increased to 16, most of the extra information still arose from words of only one or a few spikes (Fig. 3C). To understand what features of the stimulus are represented by different binary words, we therefore focused on words with only one spike.

Our next step was to construct response conditional ensembles (de Ruyter van Steveninck and Bialek, 1988): the distribution of stimuli that were associated with a particular neural response. We can think of these ensembles as “receptive fields” for the population word defined by the occurrence of a particular pattern of spiking and silences in a given time bin across the population. Given that we have just observed a particular word, we ask in what direction the stimulus was moving and when (relative to the time of our observation) it started moving. The distribution of these stimulus parameters is collected across all observations of the particular population response in the entire experiment, and the resulting probability density is displayed as a color map in Figure 4 for a sample of nine neurons. We see in Figure 4A that the occurrence of $n = 1$ spike in a population of $N = 9$ neurons is highly ambiguous in terms of the stimulus that elicited it. The red ring shows that there was a wide range of stimulus directions and

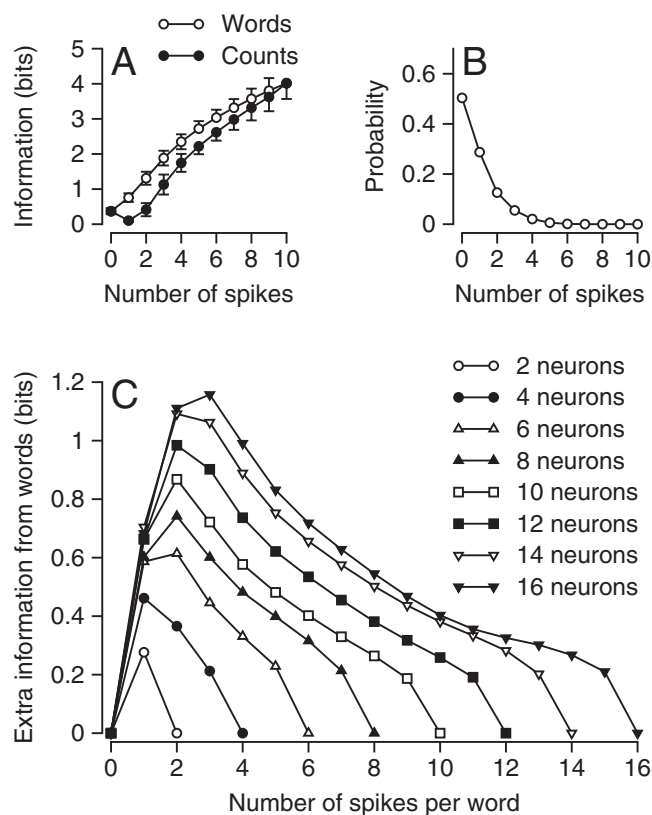


Figure 3. Added information from words analyzed separately for each spike count. **A**, Comparison of information from patterns of spiking and silences versus counts as a function of the number of spikes in the analysis window. Information from words was calculated by averaging the information from words of a given count. **B**, The probability of observing each given spike count in 100 populations of 10 MT neurons. **C**, Information from words minus information from counts is plotted separately for each spike count and each number of neurons in the analysis population. Connected sets of symbols show data for all counts in a given population size. Error bars indicate SD.

times from motion onset that had high probabilities for a count of one spike, although the neurons all had very similar direction tuning.

The event that contains one spike from nine neurons is composed of nine possible binary words, from 100000000 through 000000001, in which a single neuron in the group spikes and all others are silent. Figure 4B shows that each binary word points to a different distribution of stimuli and that each word actually represents a quite narrow range of stimulus directions and times from motion onset. Importantly, the binary words go a long way toward resolving the ambiguity between motion direction and motion onset time that is present when we look only at the pooled spike count in Figure 4A, an ambiguity that is not present in a behavior driven by the MT population response, namely smooth pursuit eye movements (Osborne et al., 2005). Notice that if the neurons really were redundant, as one might have thought from their tuning curves, then each of the events would have to point to the same distribution of stimuli, and each word would be associated with the same distribution of stimuli found whenever the spike count across the population was one. However, the small differences in tuning and response dynamics lead to regions of nonoverlap in the word-triggered receptive fields. If neuron 2 fires, for example, and all the rest are silent, then the most likely stimulus is the one that lives in the small region of nonoverlap between the receptive field of neuron 2 and the other neurons in the group. The stimulus that causes a different neuron in the group to fire and the rest of the group to remain silent could be

quite different, creating the diversity of response-conditioned stimuli shown in Figure 4B. A similar plot to Figure 4 could be constructed for any group of cells, implying that the extra information in one-spike words versus counting one spike is a general property of our MT population.

Figure 4 demonstrates that different patterns of activity across a population of MT neurons can represent different stimuli, but not necessarily that the combination of spiking in one neuron and silence in all the others is telling us anything that the labeled spike alone does not. We tested the importance of such combinations directly in Figure 5 by constructing a set of response conditional ensembles based on keeping track of the spike from a single cell and progressively discarding knowledge of silence in other cells. Although the combination of spiking in neuron 1 and silence in the rest of the population (100000000) (Fig. 5, top left color map) points to a specific, small area in the space of stimuli, specificity declines in the representation as we throw away the knowledge of silence in more and more cells. Finally, the occurrence of a spike in neuron 1 with no knowledge about the state of the other cells (1******) (Fig. 5, bottom right color map) points to a large area with tens of degrees of uncertainty about motion direction and hundreds of milliseconds of uncertainty about the time of motion onset. In the example illustrated in Figure 5, it is striking that the most uncertain large blob in the bottom right panel has almost no overlap with the original distribution of stimuli conditional on spiking in neuron 1 and silence in the others: combinations of spikes and silence not only carry more information than spike counts alone, but they also stand for very different events in the sensory input.

Returning to the quantity of information carried by combinations of spikes and silences, we found that a significant fraction of these combinations were synergistic: the information carried by the pattern of spikes and silences was larger than expected by summing the information carried by spikes and silences from the individual cells, as observed previously in the retina (Schneidman et al., 2006b). These synergistic words are the true hallmark of a combinatorial code: they cannot be read out by considering each component spike or silence independently. Whereas words generated from our sample could not code synergistically, on average, because the neurons are independent, we found that a substantial portion of the most commonly observed patterns of neural activity coded synergistically. For 10-cell groups, ~30% of all one-spike words had significant spike-silence synergy, defined as $I_{\text{synergy}} = I(\text{word}) - \sum_i I(n_i)$, where the n_i are the individual spikes and silences from each cell, labeled by i , that make up the word (see the Materials and Methods for details). The prevalence of synergy increased with N : >60% of 16-cell, one-spike words were synergistic ($I_{\text{synergy}} > 0$), suggesting that combinatorial coding is even more significant in larger cortical populations. For yet larger population sizes, we also find that many words with two spikes, for example, are synergistic, reinforcing the potential importance of combinatorial coding in large populations.

Effect of neuron–neuron correlations

So far, our discussion of the population responses in MT has assumed that the cells respond independently to sensory inputs, meaning that correlations between neurons are purely stimulus induced. We have ignored correlations generated by the brain itself, not only for simplicity but also to give the classical model of averaging over multiple redundant cells the greatest chance to succeed. We found that the diversity of temporal dynamics in neuron responses opens the possibility for a form of combinatorial coding, even among independent neurons. We now ask

whether trial-by-trial correlations among neural responses alter the potential utility of a combinatorial code.

As explained in Materials and Methods, we constructed correlated populations by fixing the mean strength of the pairwise correlations among cells, but otherwise we left the responses as random as possible, so as not to build in any structure that might artificially enhance the opportunities for combinatorial coding. As before, the time-dependent firing rates of the individual cells were matched to what we observed experimentally. We then computed the statistics of the model population responses with different levels of mean correlation and examined the information content of these responses, as for the uncorrelated populations above.

In Figure 6, we illustrate the impact of correlations on the information encoded by populations of $N = 10$ neurons. As expected from previous work (Johnson, 1980; Britten et al., 1992; Seung and Sompolinsky, 1993; Abbott, 1994; Zohary et al., 1994; Shadlen et al., 1996; Abbott and Dayan, 1999), the information available from counting spikes is reduced when we add positive correlations among cells because it increases the trial-by-trial variance of the total spike count. In contrast, negative correlations reduce the count variance and enhance information transmission. For coding based on patterns of spiking and silence across the population, small positive correlations also cause a slight drop in information that reverses as correlations become stronger, increasing the advantage of the combinatorial code over the spike count code at high levels of correlation. Across correlation levels, the extra information from a code based on words versus counts is greater than or approximately equal to that found in the independent population. Thus, the opportunities for combinatorial coding are robust across a wide range of correlation strengths, including those observed experimentally, which are usually in the range of 0.1–0.2 (van Kan et al., 1985; Gawne and Richmond, 1993; Zohary et al., 1994; Lee et al., 1998; Bair et al., 2001; Averbeck and Lee, 2003; Kohn and Smith, 2005). We conclude that combinatorial codes can exist without exotic correlations among neurons and that they are not disrupted by the modest levels of correlation consistent with available data.

Discussion

In real life, our appreciation and use of sensory information depends on interpreting individual responses of many neurons, on a time scale that makes sense in relation to behavior. With the

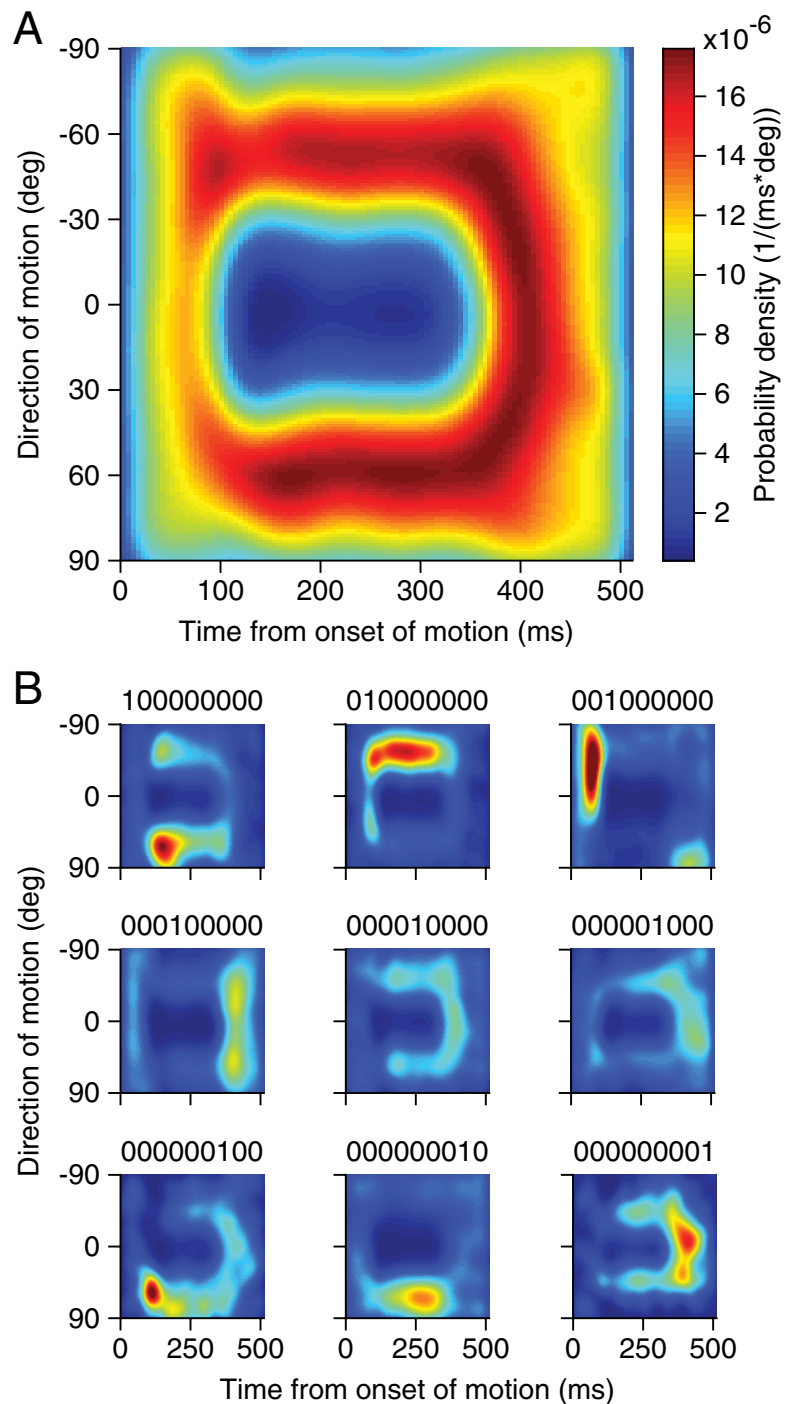


Figure 4. Response conditional stimulus ensembles for binary code words corresponding to a spike count of $n = 1$ in a population of $N = 9$ neurons. **A**, The distribution of directions of motion and delays from motion onset, $P(\theta, t - t_{\text{onset}} | n)$, given that the population of cells produced a total of one spike in window of size $\Delta\tau = 8$ ms. **B**, The same analysis, but now performed separately for each combination of spiking and silence in which one neuron emitted a spike and all the others were silent. The probabilities in **A** have been normalized so that the total integrated probability in the square is unity, with red representing the highest and dark blue representing the lowest values. The distributions in the small panels are probability densities for the stimulus conditioned on those particular one-spike events, multiplied by the probability of that event so that the average of all nine small panels yields the distribution in **A**. Graphs are based on analysis of draws from actual data in one group of nine MT neurons. $N = 9$ was chosen to allow the 3×3 presentation. deg, Degrees.

natural situation in mind, we have focused on the representation of visual motion by single trial responses across a population of neurons in cortical area MT, guided by the role that these responses play in driving smooth pursuit eye movements. Pursuit

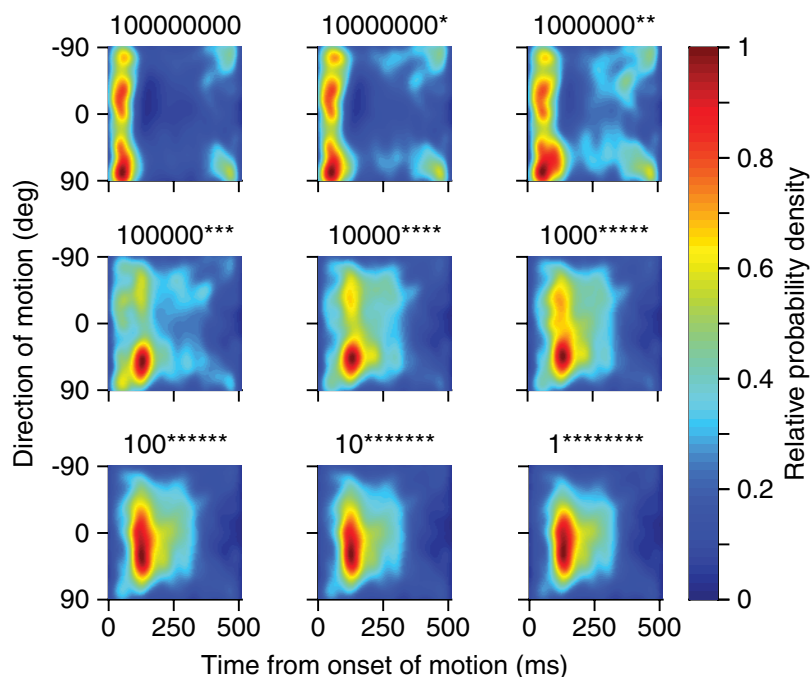


Figure 5. Contributions of silences in other neurons to the distribution of stimuli conditional on a spike in one neuron. Each pixel indicates the relative probability of a given direction of target motion at a given time after the onset of target motion, given a particular word of spiking and silences across a population of nine MT neurons. Probability densities are normalized to the peak, indicated by dark red, for each pattern. The string above each color map indicates the word that was used to create each response conditional ensemble, in which a 1 or 0 indicates the presence or absence of a spike in a neuron and an asterisk indicates a wildcard so that an interval was included in the average whether a spike was present or absent. Additional analysis revealed that the 100000000 and 1***** words contain 0.71 and 0.44 bits of information, respectively, about the stimulus. deg, Degrees.

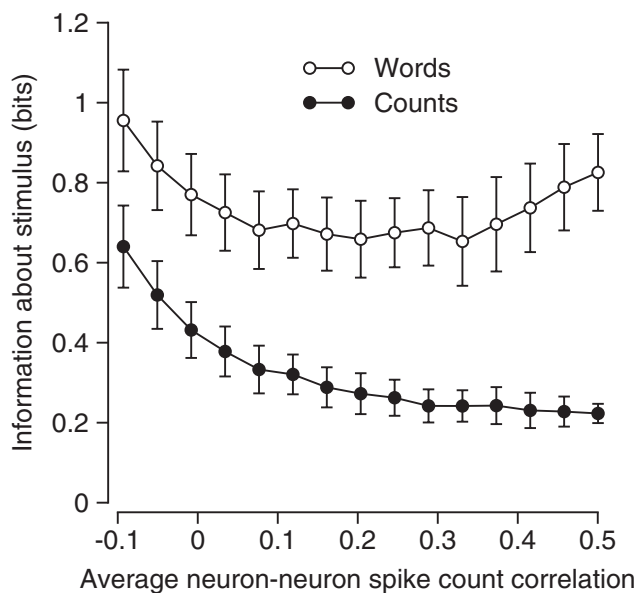


Figure 6. Impact of neuron–neuron correlations on coding based on population patterns of spikes and silence versus spike counts. The values on the x-axis are the calculated correlations between pairs of separately sampled units after setting up correlated populations using Equations 4–6. Filled and open circles show information about the stimulus from counts versus patterns of spiking and silence, respectively. Data are shown as means and SDs across 50 groups of $N = 10$ cells drawn from our experimental sample of 36 neurons.

behavior has a short latency (~ 100 ms) (Lisberger and Westbrook, 1985), reflects integration of motion signals over even shorter times scales (~ 25 ms) (Osborne and Lisberger, 2007), and has a still higher temporal precision (< 10 ms) in relation to

the visual input (Osborne et al., 2005, 2007). These behavioral constraints suggest that we should look at “the neural population response” not as an integral over hundreds of milliseconds, but rather in small windows of time, on the order of ≤ 10 ms. In such small windows, it is rare for neurons in MT to generate more than one spike: the response of an entire population of cells thus consists of a binary vector describing as 1s and 0s the pattern of spikes and silences across the members of the population. The central question about the structure of the code, then, is whether the particular combinations of 1s and 0s actually matter, or whether maximum information is available from the pooled spike counts in subpopulations of redundant cells.

The phrase “combinatorial coding” has acquired different meanings in different contexts. In recent discussions of the olfactory system, for example, the fact that single odors excite many neurons that express different receptors is taken as evidence that odors are represented by a combinatorial code (Malnic et al., 1999). In contrast, many discussions of combinatorial coding in cortex have focused on the question of whether pairs of neurons have correlations with strength or structure that depend on the particular stimulus being encoded

(Aertsen et al., 1989). We believe that there is a widely shared intuition that, in a combinatorial code, the whole message carries information that is not simply conveyed by the sum of its parts. For example, the combination of letters “th” in written English has a meaning that is not decomposable into the separate meanings of “t” and “h.” The challenge is to find a quantitative characterization of neural responses that embodies this intuition.

A minimal quantitative requirement for combinatorial coding is that the combinations of spikes and silences should carry more information than is available simply by counting the spikes. Even in the case of a population of cells with very similar feature selectivity and independent responses, we find that this is true: binary words of spiking and silence across the population contain twice as much information about the sensory input as does the pooled spike count (Fig. 1). Because we did not record from our neurons simultaneously, our results cannot arise from the correlation-based mechanism proposed by Aertsen et al. (1989), for example. Rather, we find that extra information arises from differences in response dynamics among our sample of similarly tuned cortical neurons.

It is possible to view the combinatorial code as a “labeled line” code for response dynamics in the same sense that neurons can be labeled with the preferred direction or speed to estimate those properties of the stimulus (Churchland and Lisberger, 2001). If the diversity of dynamics that we observe were to have a simple form (such as a class of cells with strongly transient responses and another class with more sustained responses), it might make sense to say that the catalog of features should be expanded to include these dynamics, perhaps generating labeled lines for constant velocities versus acceleration, which is an old idea. But the diversity that we find is not so easily categorized, and many of the

subtle differences in dynamics do not correspond in any obvious way to different preferred stimuli. Even if we could meaningfully label each individual neuron in this way, the different labeled lines would have to interact at any locus that is extracting information from the code. Therefore, it is perhaps more appropriate at this stage to focus on the binary words themselves rather than the identity of the neurons that participate in the pattern of activity. A more stringent test for combinatorial coding, then, is to look at the interactions between the elements of individual words rather than just average quantities. For example, “th” may be a prototypical example of combinatorial coding, but “qu” is almost surely redundant; averages over all pairs of letters would obscure these differences. Correspondingly, in our population of neurons, we find that different words with the same spike count stand for qualitatively different sensory stimuli (Fig. 4) and that the information carried by these words cannot be simply decomposed into contributions from spiking and silence in the individual cells (Fig. 5).

It might seem paradoxical that neurons responding independently to their sensory inputs can embody a combinatorial code. Indeed, if neurons are independent in this sense, then, on average, a population of cells must convey less information about the stimulus than expected by summing the information carried by each cell individually (Brenner et al., 2000). However, as with the examples of “th” and “qu,” this does not mean that all combinations of spiking and silence will be redundant. Indeed, in populations of 10 cells, we find that approximately one-third of all words composed from one spike and nine silences are synergistic rather than redundant, even when the neurons are conditionally independent, and that the synergistic symbols point to qualitatively different stimulus features. Importantly, from a biological perspective, the different stimuli associated with different population words serve to disambiguate stimulus identity (motion direction) from stimulus timing, and thus could play a key role in driving precise motor behaviors.

The source of the combinatorial code that we observe is not mysterious, but rather results from the known diversity of responses of MT neurons. The extra information provided by patterns of spiking and silence arises from the diversity of dynamic response properties across neurons that otherwise have very similar tuning curves for the direction of motion. The variation of response dynamics has been characterized well by previous studies (Lisberger and Movshon, 1999; Schlack et al., 2007), but discussion of the impact of response dynamics has been restricted to questions of how to extract higher derivatives of target motion from the population response. Our results suggest that cell-to-cell temporal response diversity should be considered in a broader schema for neural coding, because the full distribution of variations in the temporal response dynamics allows the population of neurons to provide substantially more information about the sensory input.

Our findings on the potential of a combinatorial code extend the results of Reich et al. (2001), obtained from samples of up to six simultaneously recorded neurons in the primary visual cortex (V1). Similarly, Schneidman et al. (2006b) found that patterns of spiking and silence across simultaneously recorded groups of up to four retinal ganglion cells carry more information than the sum of the contributions from the individual neurons. In our analysis, we have been able to show that, as one might expect, the extra information in a combinatorial code grows with the size of the neural population. Furthermore, we have been able to show that this advantage exists even when all cells in the population have exactly the same normalized tuning curves. Closely related

ideas have been explored in the theoretical work of Shamir and Sompolsky (2006), albeit not in the limit of binary responses considered here, and focusing mainly on diversity of tuning curve shapes and not on variation in dynamic response properties.

Our approach might seem to be at a disadvantage relative to Reich et al. (2001) and Schneidman et al. (2006b), because we analyzed model populations created by assembling data recorded sequentially from different neurons, whereas they analyzed populations of neurons recorded simultaneously. However, by evaluating the possibility of a combinatorial code for conditionally independent neurons, we have created the least favorable situation for finding positive results. Furthermore, by combining theory and experiment, we are able to dissect the contributions to combinatorial coding in much larger populations and to demonstrate the robustness of the essential results to plausible levels of correlation among the neural responses.

To make use of a combinatorial code, downstream neurons must combine inputs from different neurons in a way that distinguishes the different combinations of spiking and silence that we found to stand for different sensory stimuli. The study of neural sensitivity to patterns of inputs can be traced to now classical work in invertebrate systems (Segundo et al., 1963; Jacobs and Miller, 1985). More recently, it has been emphasized that simple circuits in which signals are processed along converging inhibitory and excitatory paths can evaluate the logical functions required for cells to distinguish among combinatorial code words (Schneidman et al., 2006b). Nonlinear interactions among signals converging onto the same dendrite (Polsky et al., 2004) provide additional possibilities for biologically plausible combinatorial decoding schemes in cortex.

We have presented our analysis in the concrete context provided by the coding of visual motion in area MT of the primate cortex, but the results should apply more generally. The magnitude of the extra information carried in a combinatorial code depends on the details of the neural population we are considering, but we emphasize that there is nothing extreme about the population of cells we have analyzed here. Combinatorial coding does not depend on the existence of unusual structures in the spike trains, either of single neurons or of populations. Rather, the possibility of combinatorial coding is a natural consequence of well known dynamic response properties of neurons throughout the cortex.

References

- Abbott LF (1994) Decoding neuronal firing and modeling neural networks. *Q Rev Biophys* 27:291–331.
- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Adrian ED (1926) The impulses produced by sensory nerve endings: Part I. *J Physiol* 61:49–72.
- Aertsen AM, Gerstein GL, Habib MK, Palm G (1989) Dynamics of neuronal firing correlation: modulation of “effective connectivity.” *J Neurophysiol* 61:900–917.
- Albright TD, Desimone R, Gross CG (1984) Columnar organization of directionally selective cells in visual area MT of the macaque. *J Neurophysiol* 51:16–31.
- Averbeck BB, Lee D (2003) Neural noise and movement-related codes in the macaque supplementary area. *J Neurosci* 23:7630–7641.
- Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J Neurosci* 21:1676–1697.
- Born RT, Groh JM, Zhao R, Lukasewycz SJ (2000) Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. *Neuron* 26:725–734.
- Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR (2000) Synergy in a neural code. *Neural Comput* 12:1531–1552.
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis

- of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12:4745–4765.
- Churchland MM, Lisberger SG (2001) Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. *J Neurosci* 21:9387–9402.
- de Ruyter van Steveninck RR, Bialek W (1988) Real-time performance of a movement sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. *Proc R Soc Lond Ser B* 234:379–341.
- Dubner R, Zeki SM (1971) Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res* 35:528–532.
- Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13:2758–2771.
- Groh JM, Born RT, Newsome WT (1997) How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *J Neurosci* 17:4312–4330.
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154.
- Jacobs GA, Miller JP (1985) Functional properties of individual neuronal branches isolated *in situ* by laser photoinactivation. *Science* 228:344–346.
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:62–79.
- Johnson KO (1980) Sensory discrimination: decision process. *J Neurophysiol* 43:1771–1792.
- Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25:3661–3673.
- Lee D, Port NL, Kruse W, Georgopoulos AP (1998) Variability and correlated noise in the discharge of neurons in motor parietal areas of the primate cortex. *J Neurosci* 18:1161–1170.
- Lisberger SG, Movshon JA (1999) Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *J Neurosci* 19:2224–2246.
- Lisberger SG, Westbrook LE (1985) Properties of visual inputs that initiate horizontal smooth pursuit eye movements in monkeys. *J Neurosci* 5:1662–1673.
- Malnic B, Hirono J, Sato T, Buck LB (1999) Combinatorial receptor codes for odors. *Cell* 96:713–723.
- Maunsell JH, van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49:1127–1147.
- Montani F, Kohn A, Smith MA, Schultz SR (2007) The role of correlations in direction and contrast coding in the primary visual cortex. *J Neurosci* 27:2338–2348.
- Mountcastle VB (1957) Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J Neurophysiol* 20:408–434.
- Mountcastle VB (1997) The columnar organization of the neocortex. *Brain* 120:701–722.
- Narayanan NS, Kimchi EY, Laubach M (2005) Redundancy and synergy of neuronal ensembles in motor cortex. *J Neurosci* 25:4207–4216.
- Newsome WT, Wurtz RH, Dürsteler MR, Mikami A (1985) Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *J Neurosci* 5:825–840.
- Osborne LC, Lisberger SG (2007) Spatial and temporal integration of stochastic motion direction signals in primate pursuit eye movements. *Soc Neurosci Abstr* 33:715.24.
- Osborne LC, Bialek W, Lisberger SG (2004) Time course of information about motion direction in visual area MT of macaque monkeys. *J Neurosci* 24:3210–3222.
- Osborne LC, Lisberger SG, Bialek W (2005) A sensory source for motor variation. *Nature* 437:412–416.
- Osborne LC, Hohl SS, Bialek W, Lisberger SG (2007) Time course of precision in smooth pursuit eye movements of monkeys. *J Neurosci* 27:2987–2998.
- Panzeri S, Schultz SR, Treves A, Rolls ET (1999) Correlations and the encoding of information in the nervous system. *Proc R Soc Lond B Biol Sci* 266:1001–1012.
- Perkel DH, Bullock TH (1968) Neural coding. *Neurosci Res Prog Sum* 3:405–527.
- Polsky A, Mel BW, Schiller J (2004) Computational subunits in thin dendrites of pyramidal cells. *Nat Neurosci* 7:567–569.
- Reich DS, Mechler F, Victor JD (2001) Independent and redundant information in nearby cortical neurons. *Science* 294:2566–2568.
- Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: exploring the neural code*. Cambridge, MA: MIT Press.
- Schlack A, Krekelberg B, Albright TD (2007) Recent history of stimulus speeds affects the speed tuning of neurons in area MT. *J Neurosci* 27:11009–11018.
- Schneidman E, Still S, Berry II MJ, Bialek W (2003) Network information and connected correlations. *Phys Rev Lett* 91:238701–231704.
- Schneidman E, Berry MJ 2nd, Segev R, Bialek W (2006a) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
- Schneidman E, Puchalla JL, Harris RA, Bialek W, Berry MJ 2nd (2006b) Synergy from silence in a combinatorial neural code. *q-bio.NC/0607017*.
- Segundo JP, Moore GP, Stensaas LJ, Bullock TH (1963) Sensitivity of neurons in *Aplysia* to temporal pattern of arriving impulses. *J Exp Biol* 40:643–667.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753.
- Shadlen MN, Britten KH, Newsome WT, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* 16:1486–1510.
- Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. *Neural Comput* 18:1951–1986.
- Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.
- van Kan PLE, Scobey RP, Gabor AJ (1985) Response covariance in cat visual cortex. *Exp Brain Res* 60:559–563.
- Van Rullen R, Thorpe SJ (2001) Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput* 13:1255–1283.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.