Decoding neuronal spike trains: How important are correlations?

Sheila Nirenberg* and Peter E. Latham

Department of Neurobiology, University of California, Los Angeles, CA 90095-1763

Communicated by Charles R. Gallistel, Rutgers, The State University of New Jersey at New Brunswick, Piscataway, NJ, April 1, 2003 (received for review February 17, 2003)

It has been known for >30 years that neuronal spike trains exhibit correlations, that is, the occurrence of a spike at one time is not independent of the occurrence of spikes at other times, both within spike trains from single neurons and across spike trains from multiple neurons. The presence of these correlations has led to the proposal that they might form a key element of the neural code. Specifically, they might act as an extra channel for information, carrying messages about events in the outside world that are not carried by other aspects of the spike trains, such as firing rate. Currently, there is no general consensus about whether this proposal applies to real spike trains in the nervous system. This is largely because it has been hard to separate information carried in correlations from that not carried in correlations. Here we propose a framework for performing this separation. Specifically, we derive an information-theoretic cost function that measures how much harder it is to decode neuronal responses when correlations are ignored than when they are taken into account. This cost function can be readily applied to real neuronal data.

Ever since Adrian and Zotterman observed that the firing rate of peripheral touch receptors coded for the pressure applied to a patch of skin (1), neuroscientists have been trying to crack the neural code, that is, to understand the relationship between neuronal activity and events in the outside world. For much of that time, the working hypothesis was that information is carried by firing rate. More recently it has been proposed that firing rate is not the whole story: Information might also be carried in spike patterns, both within spike trains from single neurons (2–6) and across spike trains from multiple neurons (7–9).

One aspect of this proposal, an aspect that has led to a great deal of debate, is that correlations in spike patterns may be of particular importance (7–10, 12). It has been known for many years that spike trains contain correlations; that is, the presence of a spike at one time is not independent of the presence of spikes at other times. These correlations exist not just within spike trains but across them as well, with the most common example being synchronous spikes across pairs of cells (13–16). What has led to the debate is the suggestion that these correlations might form a key aspect of the code. The idea is that they might serve as an extra information channel, conveying messages not carried elsewhere in the spike trains.

How might the correlations do this? An example, using synchronous spikes, is shown in Fig. 1*a*. In this example there are two stimuli, A and B, and two neurons. When stimulus A is presented, the two neurons produce five spikes on average. When stimulus B is presented, they also produce five spikes on average. What is different, though, is the correlational structure of the responses: When stimulus A is presented, the two neurons tend to produce few synchronous spikes, whereas when stimulus B is presented, they produce many. Thus, the difference in the degree of synchrony is essential; one cannot tell the stimuli apart without it.

The alternative is that the correlations do not carry extra information but instead carry only information that is redundant to what is carried elsewhere in the spike trains (e.g., in the spike count). An example of this, in which the number of synchronous spikes depends on spike count but not directly on the stimulus, is shown Fig. 1*b*. As in Fig. 1*a*, there are two stimuli, A and B, and two neurons. When stimulus A is presented, the two neurons produce five spikes on average. When stimulus B is presented, they produce many more spikes, 10 on average. The number of synchronous spikes is also higher with stimulus B, but here it is just a consequence of the larger spike count. Thus, although one could use the difference in the number of synchronous spikes to tell the stimuli apart, one does not have to. One could just use the difference in the number of spikes.

Why has the debate about whether correlations are important been hard to resolve? The main reason is that real data are rarely as unambiguous as they are in these two cases. In more realistic situations, there are a large number of stimuli, the correlations are often more complicated than those associated with synchronous spikes (4, 6), and both firing rate and correlations vary with time. Consequently, information in correlations is often tied to information not in correlations in subtle, hard-to-disentangle ways.

Here we describe an approach for assessing the role of correlations. The approach is to ignore them (by treating, for example, two cells in a pair as independent, a notion that will be made explicit below) and ask how much this affects our ability to determine what the stimulus is. The approach is general enough that it can be used for correlations of arbitrary complexity, both across spike trains and within spike trains, thus it goes beyond simply assessing the role of correlations found in synchronous spikes. Moreover, it can be applied to neuronal data in a straightforward manner.

The General Problem of Decoding

We address the question of whether correlations are important in the context of decoding. By "decoding" we mean building a dictionary that translates responses into stimuli. To build this dictionary, we use a general probabilistic approach: We first determine the stimulus-to-response relationship and then use that, via Bayes' theorem, to find the inverse, the response-to-stimulus relationship. Experimentally, we present stimuli over and over and obtain a histogram of responses for each stimulus. We then use these histograms to estimate the probability that a particular response occurred given that a particular stimulus occurred. This quantity is denoted $p(\mathbf{r}|s)$, where $\mathbf{r} \equiv (r_1, r_2, \ldots, r_n)$ is a set of n neuronal responses, and s is the stimulus. (Here the different r_i could be responses from different neurons, from the same neuron in different time bins, or some combination of the two.) Once we know $p(\mathbf{r}|s)$, we apply Bayes' theorem to derive the inverse, the probability that a particular stimulus occurred given that a particular response occurred. This quantity is denoted $p(s|\mathbf{r})$ and is given by

$$p(s|\mathbf{r}) = \frac{p(\mathbf{r}|s)p(s)}{p(\mathbf{r})}, \qquad [1]$$

*To whom correspondence should be addressed at: Department of Neurobiology, University of California, 10833 Le Conte Avenue, Los Angeles, CA 90095-1763. E-mail: sheilan@ucla.edu.

 www.pnas.org/cgi/doi/10.1073/pnas.1131895100

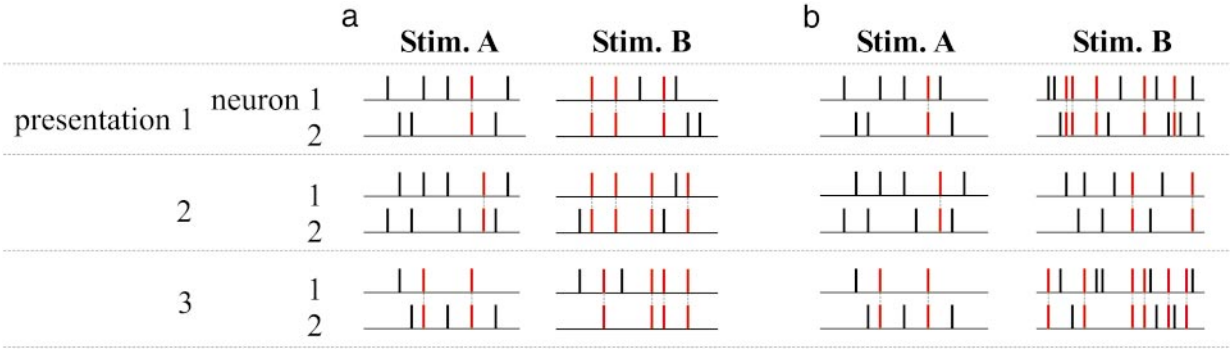

Fig. 1. Two scenarios: one in which correlations are critical for determining what the stimulus is, and one in which they are not. Shown are the results of a hypothetical experiment in which two stimuli, A and B, are presented several times, and the responses of two neurons are recorded. Synchronous spikes are marked in red and linked by a dashed gray line. (*a*) Both stimuli produce five spikes on average, but the number of synchronous spikes is higher for stimulus B than for A. Thus, knowledge of the difference in the degree of synchrony is needed to distinguish the stimuli. (*b*) Stimulus B produces, on average, more spikes than stimulus A (10 versus 5). The number of synchronous spikes is also higher for stimulus B but only because it is a function of spike count. Here, knowledge of the difference in the degree of synchrony is not needed to distinguish the stimuli; one can just use the difference in the spike count.

where $p(s)$ is the probability that stimulus s occurred, and $p(\mathbf{r}) \equiv \Sigma_s\, p(\mathbf{r}|s)p(s)$ is the probability that response $\mathbf{r}$ occurred. The quantity $p(s|\mathbf{r})$ is our dictionary. A simple example showing the construction of such a dictionary for two stimuli and one neuron is shown in Fig. 2.

Defining "Correlated"

There are two types of correlations in the literature. One is called "noise correlation" (17) and is defined as follows: neuronal responses are noise-correlated if and only if

$$p(r_1, r_2, \ldots, r_n|s) \neq \prod_{i=1}^{n} p(r_i|s). \qquad [2]$$

The second type is called "signal correlation" (17) and differs from noise correlation in that it incorporates an average over stimuli. Specifically, responses are signal-correlated if and only if

$$p(r_1, r_2, \ldots, r_n) \neq \prod_{i=1}^{n} p(r_i).$$

The following example illustrates the difference. Suppose we present a flash of light while recording from two ON-type retinal ganglion cells that lie far apart on the retina (far enough that their receptive fields do not overlap). Because the cells are both ON-type, they will both fire when the light is flashed on. This similarity in their response is an example of signal correlations, and its role in neural coding is obvious and undisputed. If, on the other hand, the two ON-type cells are close enough to receive common input from presynaptic cells (e.g., common photoreceptors, amacrine cells, etc.), then they would exhibit correlations above the signal correlations. These extra correlations are noise correlations, the ones whose function have become the subject of debate. It is these that we focus on in this article. Thus, for the remainder of this article, when we refer to "correlated" we mean "noise-correlated."

A General Approach for Assessing the Importance of Correlations for Decoding

Our approach to assessing the importance of correlations is to ignore them and determine how this affects our ability to decode responses. We ignore them by treating the responses as though they were independent; formally, we replace $p(\mathbf{r}|s)$ with the independent distribution, $\Pi_i\, p(r_i|s)$, the latter denoted $p_{\text{ind}}(\mathbf{r}|s)$.

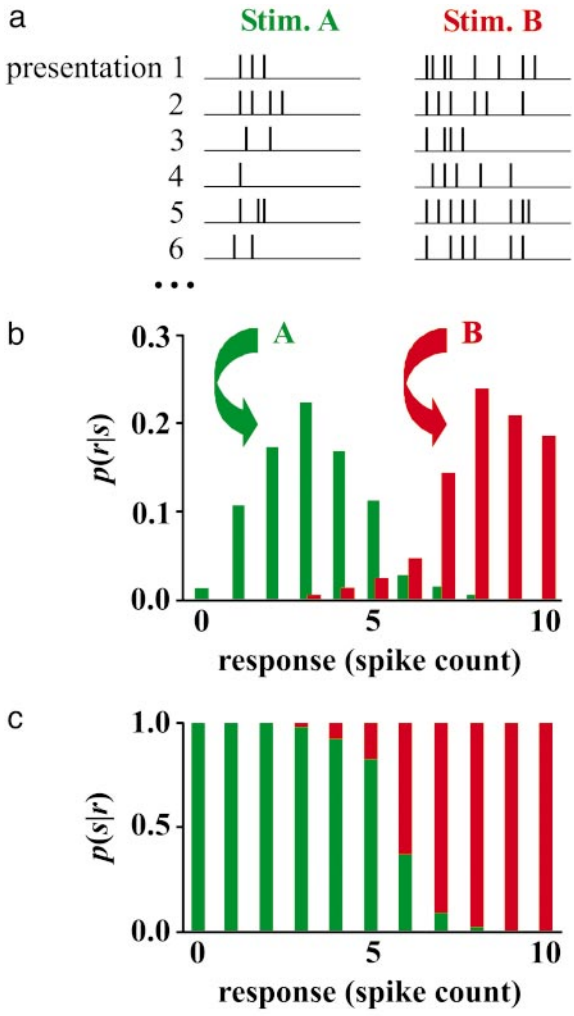

Fig. 2. Determining $p(s|\mathbf{r})$, the probability that a particular stimulus occurred given that a particular response occurred. Here the response, r, is the spike count of one neuron. (*a*) Outcome of six presentations for stimuli A (*Left*) and B (*Right*). (*b*) Histogram showing the probability that a particular response occurred given that a particular stimulus occurred. (*c*) Probability that each of the two stimuli occurred given that a particular response occurred, constructed using Eq. **1**. Response is given on the horizontal axis; the lengths of the green and red bars are the probabilities that stimuli A and B occurred, respectively, given a response.

By definition, making such a replacement eliminates all correlations (see Eq. **2**). To see whether this replacement affects our ability to determine stimuli from responses, we insert $p_{\text{ind}}(\mathbf{r}|s)$ into Bayes' theorem, Eq. **1**. This gives us the "independent" estimate of the probability of stimuli given responses, denoted $p_{\text{ind}}(s|\mathbf{r})$,

$$p_{\text{ind}}(s|\mathbf{r}) = \frac{p_{\text{ind}}(\mathbf{r}|s)p(s)}{p_{\text{ind}}(\mathbf{r})}, \qquad [3]$$

where $p_{\text{ind}}(\mathbf{r}) \equiv \Sigma_s\, p_{\text{ind}}(\mathbf{r}|s)p(s)$ is the independent total response distribution.

This is the main idea behind our approach: We approximate $p(s|\mathbf{r})$, the true distribution of stimuli given responses, with $p_{\text{ind}}(s|\mathbf{r})$, the distribution one would derive in the absence of knowledge of correlations, and ask whether it matters. If $p_{\text{ind}}(s|\mathbf{r}) = p(s|\mathbf{r})$, then we know that correlations are not important for decoding. If $p_{\text{ind}}(s|\mathbf{r}) \neq p(s|\mathbf{r})$, then we know that they are; there will be a cost associated with ignoring them. Our aim now is to determine what that cost is.

Quantifying the Cost of Ignoring Correlations

Ideally, one would like to determine the behaviorally relevant cost of ignoring correlations, that is, the cost to an animal of using $p_{\text{ind}}(s|\mathbf{r})$ rather than $p(s|\mathbf{r})$ to decode responses. However, such a cost is hard to derive experimentally. Moreover, in many cases it does not exist in an absolute sense, because cost is almost always context-dependent. For example, when approaching an intersection in a car, the behaviorally relevant quantity is the color of the traffic lights, whereas when reading a digital clock, the relevant quantity is the shape of the lights.

For these reasons, we used an information-theoretic approach. The advantage is that it is general and context-independent; the disadvantage is that it may misestimate the behavioral relevance of some stimuli. The idea behind this approach is as follows: Neuronal responses provide information about the stimuli that produce them. To extract all the information they carry, one must have knowledge of the true distribution of stimuli given responses, $p(s|\mathbf{r})$. Knowledge of only the independent distribution, $p_{\text{ind}}(s|\mathbf{r})$, leads to a reduction in the amount of information one can extract. Thus, there is an information-theoretic cost to ignoring correlations.

To quantify the cost, we begin with the expression for mutual information, that is, the information the responses provide about the stimuli (18),

$$I(s;\,\mathbf{r}) = -\sum_s p(s)\log_2 p(s) + \sum_{s,\mathbf{r}} p(s,\,\mathbf{r})\log_2 p(s|\mathbf{r}), \qquad [4]$$

where $p(s,\,\mathbf{r})$ is the joint stimulus–response distribution.

Classically, information theory is a theory about true probability distributions. To understand how to incorporate approximate distributions, in particular $p_{\text{ind}}(s|\mathbf{r})$, into an information-theoretic framework, we recast mutual information using the yes/no-question formulation given by Cover and Thomas (19). With this formulation, mutual information is described in the context of a guessing game in which stimuli are drawn from a distribution, and one has to determine what they are by asking yes/no questions. In this context, the first term in Eq. **4** is the average number of yes/no questions one would have to ask to determine the stimuli without the benefit of observing neuronal responses, and the second term is the average number of yes/no questions one would have to ask to determine the stimuli with the benefit of observing neuronal responses. For both terms, the assumption is that the optimal question-asking strategy is used.

For example, consider a stimulus set containing 16 items, all of which occur with equal probability. A stimulus is drawn, and we must determine what it is. Taking the optimal question-asking strategy, the number of yes/no questions we would have to ask is four. This is because the optimal strategy is to divide the possibilities in half with each question. Thus, our first question would be "is it stimulus 1–8?" If the answer is no, we then would ask "is it stimulus 9–12?" etc., until we arrive at the correct answer. Because we use this strategy of dividing the possibilities in half or, stated more generally, of dividing the total stimulus probability in half, the number of yes/no questions needed to determine a stimulus with probability $p(s)$ is $[-\log_2 p(s)]$.[†] If the game is played repeatedly, then the number of yes/no questions needed to determine the stimuli on average is $-\Sigma_s\, p(s)\log_2 p(s)$.

Now consider the situation where we have the benefit of observing neuronal responses, which changes the probability of stimuli from $p(s)$ to $p(s|\mathbf{r})$. Consequently, the number of yes/no questions needed to determine the stimulus changes from $[-\log_2 p(s)]$ to $[-\log_2 p(s|\mathbf{r})]$. For example, if a response told us that the stimulus was not stimulus 1–8, then the number of yes/no questions we would have to ask would change from four to three. On average, then, we have

$$\langle \text{number of yes/no questions}|\text{responses}\rangle = -\sum_{s,\mathbf{r}} p(s,\,\mathbf{r})\log_2 p(s|\mathbf{r}). \qquad [5]$$

The key idea to be extracted from this yes/no-question formulation is that the stimulus probabilities induce a question-asking strategy. When we know what the stimulus probabilities are, as was the case above, we are able to take the optimal question-asking strategy and solve the guessing game with the fewest possible questions. If, however, we do not know the stimulus probabilities, then we are left to take a suboptimal strategy (see Fig. 3 for an example).

This is the situation that arises when we approximate $p(s|\mathbf{r})$ with $p_{\text{ind}}(s|\mathbf{r})$. When we make this approximation, we have to construct our question-asking strategy in a suboptimal manner and thus will likely need to ask more questions. Specifically, when we use $p_{\text{ind}}(s|\mathbf{r})$, the number of yes/no questions we need to ask to guess a stimulus given a response is $[-\log_2 p_{\text{ind}}(s|\mathbf{r})]$. The average number is thus

$$\langle \text{number of yes/no questions}|\text{responses}\rangle_{\text{ind}} = -\sum_{s,\mathbf{r}} p(s,\,\mathbf{r})\log_2 p_{\text{ind}}(s|\mathbf{r}), \qquad [6]$$

where the subscript "ind" on the left-hand side indicates that the independent distribution is being used rather than the true one. Note that to derive Eq. **6** we averaged over the true distribution, $p(s,\,\mathbf{r})$, rather than the independent one, $p_{\text{ind}}(s,\,\mathbf{r}) \equiv p_{\text{ind}}(s|\mathbf{r})p_{\text{ind}}(\mathbf{r})$. This is because the question-asking strategy one uses has no effect on the probability of a particular stimulus–response pair occurring.

We now arrive at our cost function, the cost of using $p_{\text{ind}}(s|\mathbf{r})$ rather than $p(s|\mathbf{r})$. The cost is the extra number of yes/no questions we would have to ask to determine the stimuli; it is just the difference between the right-hand sides of Eqs. **5** and **6**. Denoting this difference ΔI, we find that

$$\Delta I = \sum_{s,\mathbf{r}} p(s,\,\mathbf{r})\log_2 \frac{p(s|\mathbf{r})}{p_{\text{ind}}(s|\mathbf{r})}. \qquad [7]$$

Note that ΔI is the conditional relative entropy, or average Kullback–Leibler distance, between $p(s|\mathbf{r})$ and $p_{\text{ind}}(s|\mathbf{r})$ (19). It is not itself a mutual information, but it is measured in bits (19)

[†]For completeness, the number of yes/no questions should be $[-\log_2 p(s)]$ rounded up to the nearest integer. However, as shown in ref. 19, the effect of rounding is negligible.

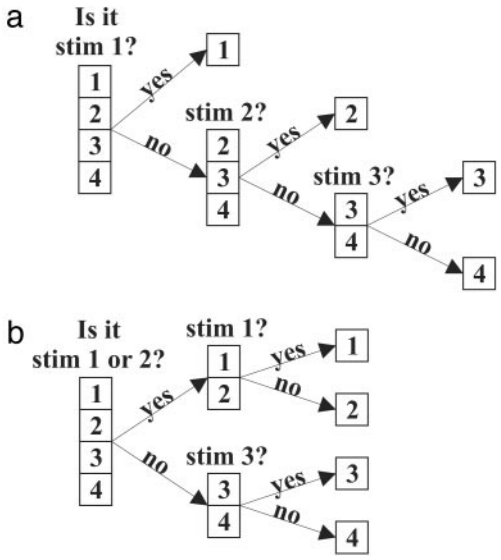

Fig. 3. The stimulus distribution induces an optimal question-asking strategy. (*a*) Consider a distribution with four stimuli, numbered 1–4, and occurring with probabilities 1/2, 1/4, 1/8, and 1/8, respectively. A stimulus is drawn repeatedly from this distribution, and each time we must ask yes/no questions to determine what it is. The optimal question-asking strategy is to divide the probability in half with each question. Thus for this distribution, our first question is ''is it stimulus 1?'' If the answer is ''no,'' we then ask ''is it stimulus 2?'', etc., until we arrive at the correct answer. With this strategy, stimulus 1 is guessed in one question, stimulus 2 in two, and stimuli 3 and 4 in three. On average, the number of questions we will ask to determine the stimuli is (1/2)1 + (1/4)2 + (1/8)3 + (1/8)3 = 1 3/4. (*b*) In this case, we do not know the stimulus probabilities and so make the assumption that they occur with equal probability. This wrong assumption would cause us to use a suboptimal question-asking strategy. Our first question would be ''is it stimulus 1 or 2?'' because that question would divide the assumed probabilities in half. However, that question does not divide the true probabilities in half. Continuing with this strategy, we see that all stimuli are guessed in two questions. The average number of questions is thus (1/2)2 + (1/4)2 + (1/8)2 + (1/8)2 = 2, which is greater than the average number shown in *a*.

(which are essentially synonymous with yes/no questions) and so can be compared directly to the mutual information, I, between stimuli and responses. Thus, we interpret $\Delta I/I$ as the relative cost of ignoring correlations.

It can be shown that ΔI is nonnegative (19), and that it reaches its minimum, zero, only when $p_{\text{ind}}(s|\mathbf{r}) = p(s|\mathbf{r})$, that is, only when correlations are not important. It can also be shown that ΔI reduces to a standard measure for comparing neural codes (ΔI is the difference in mutual information between, for example, codes with large bins and codes with small bins). The latter property is demonstrated in *The Information Difference Between Two Codes Is Equal to* ΔI, which is published as supporting information on the PNAS web site, www.pnas.org.

The expression for ΔI given in Eq. **7** was first derived by Nirenberg *et al.* (20). More recently, Pola *et al.* (21) derived an identical expression based on slightly different considerations; they called it $I_{\text{cor}-\text{dep}}$ rather than ΔI.

Note that $\Delta I = 0$ can mean one of two things: Either correlations do not exist, and thus are obviously not important for decoding, or they do exist but are not needed for the decoding process. The latter is a key point: $p_{\text{ind}}(s|\mathbf{r})$ can equal $p(s|\mathbf{r})$ even though $p_{\text{ind}}(\mathbf{r}|s) \neq p(\mathbf{r}|s)$. This can happen if correlations are not stimulus-dependent, as in Fig. 1*b* (see also *Existence Versus Importance of Correlations*, which is published as supporting information on the PNAS web site). Consequently, cost functions that simply measure the strength of correlations, for example, cost functions that measure the distance between $p_{\text{ind}}(\mathbf{r}|s)$ and $p(\mathbf{r}|s)$, tell us whether correlations exist but not whether they are important for decoding.

Other Measures of the Importance of Correlations

Historically, the primary method for assessing the importance of correlations has been to look for stimulus-dependent changes in cross-correlograms (22–26). There are, however, two problems with this approach. First, firing rate can have a large effect on the shape of a cross-correlogram, making it difficult to separate information carried in firing rate from information carried in correlations. Second, methods based on cross-correlograms are sensitive to only one kind of correlational structure, synchronous or near-synchronous spikes, and may miss others. In particular, they cannot be used to evaluate the role of correlations within spike trains of single neurons, nor can they be used for correlations among complex spike patterns across multiple neurons.

In principle, both of these problems, lack of quantifiability and lack of sensitivity to complex patterns, can be overcome by using information-theoretic measures. Developing such measures, however, has been harder than expected. In particular, two that have appeared in the literature, shuffled information (27, 28) and synergy/redundancy (6, 29), seem intuitive but, in fact, turn out to be confounded when used to assess the role of correlations for decoding. Below we review them briefly.

The idea of shuffling neuronal responses to test a broad range of correlational structures was proposed in general by us (30) and later refined by Panzeri *et al.* (27, 28). The idea is as follows: present a set of stimuli while recording simultaneously from two neurons and estimate the information between stimuli and responses; then, remove the correlations by shuffling the responses such that the neurons see the same stimuli but at different times, and again estimate the information between stimuli and (the now-uncorrelated) responses. With this method, the quantity used to assess the importance of correlations is the difference between these two informations. The difference, denoted $\Delta I_{\text{shuffled}}$, is given by

$$\Delta I_{\text{shuffled}} = I(s; \mathbf{r}) - I_{\text{shuffled}}(s; \mathbf{r}),$$

where $I(s;\mathbf{r})$ is the mutual information between the stimulus and the correlated responses (Eq. **4**), and $I_{\text{shuffled}}(s; \mathbf{r})$ is the mutual information between the stimulus and the uncorrelated responses. The latter quantity is defined by

$$I_{\text{shuffled}}(s; \mathbf{r}) \equiv -\sum_{s} p(s)\log_2 p(s) + \sum_{s,\mathbf{r}} p_{\text{ind}}(s, \mathbf{r})\log_2 p_{\text{ind}}(s|\mathbf{r}).$$

Because shuffling removes correlations, one might expect that if correlations are not important for decoding, $\Delta I_{\text{shuffled}}$ will be zero, whereas if they are, $\Delta I_{\text{shuffled}}$ will be nonzero. This, however, turns out not to be the case. The reason is that correlations can either increase or decrease information (31), so it is possible for some parts of the stimulus–response space to exhibit correlations that increase $\Delta I_{\text{shuffled}}$ while other parts exhibit correlations that decrease it (e.g., for a given pair of cells, $\Delta I_{\text{shuffled}}$ might increase for moving stimuli but decrease for colored stimuli). Notably, the increase and decrease can exactly cancel, making $\Delta I_{\text{shuffled}}$ zero in cases where correlations actually are important for decoding. Less obviously, $\Delta I_{\text{shuffled}}$ can be nonzero when correlations are not important. As a result, no reliable relationship exists between $\Delta I_{\text{shuffled}}$ and the importance of correlations. Specifically, we show in *Evaluation of Other Measures for Decoding*, which is published as supporting information on the PNAS web site (see especially Fig. 5 and Table 1, which are also published as supporting information on the PNAS web site), that $\Delta I_{\text{shuffled}}$ can be positive, negative, or zero both when correlations are not important for decoding (when ignor-

ing them has no effect on our estimate of the stimulus) and when they are important (when ignoring them does have an effect).

These conclusions are not strictly theoretical; an example in which $\Delta I_{\text{shuffled}}$ is large even though correlations are not important for decoding has been seen in real data. Panzeri *et al.* (32) examined the information from pairs of neurons in rat barrel cortex and found that, averaged over 39 cell pairs, $\Delta I_{\text{shuffled}}$/I was $\approx$20%, even though essentially no information was lost by ignoring correlations ($\Delta I/I < 2\%$). This means, as pointed out by the authors, that a decoder could completely ignore correlations and still recover almost all the information in the spike trains.

Although $\Delta I_{\text{shuffled}}$ cannot be used to assess directly the role of correlations in decoding information, it can be used to provide information about how correlations affect the transformation from stimulus to response (21, 31, 33). In particular, the quantity $\Delta I_{\text{shuffled}} - \Delta I$ is especially useful for understanding the relation between signal and noise correlations (21).

A second information-theoretic measure aimed at quantifying the role of correlations is synergy/redundancy. This measure is the total information that neuronal responses provide about a set of stimuli minus the information provided by the individual responses taken separately. Specifically, if, as above, $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ where the r_i are responses from individual neurons or individual time bins, then the synergy/redundancy measure, denoted $\Delta I_{\text{synergy}}$, is given (6) by

$$\Delta I_{\text{synergy}} = I(s; \mathbf{r}) - \sum_i I(s; r_i).$$

For $\Delta I_{\text{synergy}}$ to be positive, there needs to be some cooperative coding among the responses. Such coding is exhibited by, for example, two neurons whose responses to two stimuli are the following: the neurons produce identical responses when one of the stimuli is present and different responses when the other is present, and separately are uncorrelated with the stimuli. Neither response alone tells us anything at all about the stimuli, so $I(s; r_1) = I(s; r_2) = 0$. However, if both responses are observed simultaneously, the stimulus can be uniquely decoded, which means that $I(s; r_1, r_2) = 1$ (assuming the stimuli appear with equal probability). Thus, $\Delta I_{\text{synergy}} = 1$. Here, correlations are critical for decoding the responses; this is reflected in ΔI, which is also equal to 1.

The idea that cooperative coding (another name for coding with correlations) is necessary for $\Delta I_{\text{synergy}}$ to be positive has led to the view that the converse is true, that if $\Delta I_{\text{synergy}}$ is positive, then correlations are important for decoding. However, like $\Delta I_{\text{shuffled}}$, $\Delta I_{\text{synergy}}$ can be positive when correlations are not important, i.e., when $p_{\text{ind}}(s|\mathbf{r}) = p(s|\mathbf{r})$. In addition, $\Delta I_{\text{synergy}}$ has the same potential for cancellation effects as $\Delta I_{\text{shuffled}}$: It is possible for some parts of the stimulus–response space to exhibit correlations that increase $\Delta I_{\text{synergy}}$, while other parts exhibit correlations that decrease it. This makes values of $\Delta I_{\text{synergy}}$ near zero hard to interpret. Consequently, although $\Delta I_{\text{synergy}}$ may be useful for evaluating some aspects of the neural code [see, for example, Brenner *et al.* (6)], it cannot by itself be used to evaluate the importance of correlations for decoding. Specifically, we show in *Evaluation of Other Measures for Decoding* (see especially Fig. 6 and Table 1, which are published as supporting information on the PNAS web site) that $\Delta I_{\text{synergy}}$ can be positive, negative, or zero both when correlations are not important for decoding (when ignoring them has no effect on our estimate of the stimulus) and when they are important (when ignoring them does have an effect).

Other information-theoretic measures have been more successful at assessing the importance of correlations. Dan *et al.* (34) and Oram *et al.* (35) assessed the importance of synchronous spikes by computing mutual information with and without them. They did this by dividing responses from two neurons into time bins and either counting or not counting the number of synchronous spikes in each bin when computing mutual information (Fig. 4, which is published as supporting information on the PNAS web site). Interestingly, the difference in information computed in this way turns out to equal ΔI for their specific code (see *The Information Difference Between Two Codes Is Equal to* ΔI*)*.

This method is ideally suited for assessing the role of synchronous spikes, provided the responses change slowly enough that reasonably large bins can be used (it cannot be used in the limit of very small bins, because the difference in information vanishes in this limit). A limitation of the method, though, is that it, like the cross-correlogram method, captures the effects of only one kind of correlations, synchronous spikes across multiple neurons.

Panzeri and colleagues (36–38) took a different approach. They computed mutual information in the limit of small time bins and found that it broke naturally into four terms. Two of those terms depend on correlations, in the sense that they vanish when correlations are absent. Nonzero values for those two correlation-dependent terms thus were interpreted to imply that correlations are important for transmitting information (36–38).

Their method has now been made general in the sense that it no longer requires small time bins (21). This generalized method also yields two terms that depend on correlations, which are called $I_{\text{cor-ind}}$ and $I_{\text{cor-dep}}$. The latter, $I_{\text{cor-dep}}$, is identical to ΔI, whereas the sum of the two, $I_{\text{cor-ind}} + I_{\text{cor-dep}}$, is equal to $\Delta I_{\text{shuffled}}$.

The difference between our work and that of Panzeri and colleagues is that we used a top-down rather than a bottom-up approach: We started with general arguments about the role of correlations in decoding neuronal responses and derived a cost function guaranteed to measure the importance of correlations; Panzeri *et al.* started with the expression for mutual information and separated out the terms that depended on correlations. The top-down approach allowed $I_{\text{cor-dep}}$ but not $I_{\text{cor-ind}}$ to be interpreted as measuring the importance of correlations for decoding.

Finally, Wu and colleagues (39, 40) used an approach similar to ours to investigate theoretically the role of correlations in large population codes. They also ignored correlations and asked how much that affected one's ability to determine the stimuli from the responses. There were, however, two main differences between their work and ours. First, they considered, theoretically, a population of neurons coding for a single variable and assumed that the conditional response distribution was multivariate Gaussian with known covariance matrix. Second, they used Fisher information, which is related to the minimum variance of a deterministic decoder rather than ΔI to assess the importance of correlations. They asked, for two particular covariance matrices, how much the Fisher information changed if one ignored correlations, that is, how much it changed if one ignored the off-diagonal terms in the covariance matrix. Interestingly, for the covariance matrices they considered, the changes were small, meaning correlations were not very important.

Discussion

We have developed an approach for assessing the importance of correlations in decoding neuronal responses. The cost function that measures their importance, ΔI, is the extra number of yes/no questions it would take to determine a set of stimuli given responses, assuming that one had access only to the independent distribution, $\Pi_i\, p(r_i|s)$, and not the true one, $p(\mathbf{r}|s)$. It is also the Kullback–Leibler distance between $p(s|\mathbf{r})$, the stimulus distribution built with knowledge of the correlations, and $p_{\text{ind}}(s|\mathbf{r})$, the distribution built without such knowledge. Although other cost functions may be more appropriate for specific problems (39, 40), this one has three advantages. First, if ΔI is zero, then ignoring correlations will have absolutely no effect on our ability to decode responses; that is, if we were to build a decoder using

the independent distribution, it would translate responses identically to a decoder built from the true distribution. This is because $\Delta I = 0$ means $p_{\text{ind}}(s|\mathbf{r}) = p(s|\mathbf{r})$. Second, the cost function is general enough to pick up all effects of correlations (not just synchrony). Finally, the cost function can be compared directly to the mutual information, I, between stimuli and responses. This last statement allows us to interpret the ratio $\Delta I/I$ as the relative cost of ignoring correlations.

Why was it necessary to develop this framework, that is, why was it not possible to just compare the information in neuronal responses with and without correlations taken into account? The latter approach has been used successfully for comparing many different coding schemes including temporal versus rate coding (41), small versus large time bins (42–44), and labeled line versus population averages (45). The reason this does not work for correlations, however, is that assessing the importance of correlations is fundamentally different from comparing codes: In the latter, information about the responses is thrown away, and in the former, information about the response distribution is thrown away. This qualitative difference required us to make a comparison in a more general way by looking at the cost, in yes/no questions, of ignoring correlations. This cost turned out to equal the information difference when comparing two codes (see *The Information Difference Between Two Codes Is Equal to* ΔI), thus in more standard situations it reduces to the appropriate measure.

Determining the role of correlations has important practical implications. If correlations are not important, then the problem of building a decoder is greatly simplified, because one needs to measure only single-neuron or single time-bin response distributions; if correlations are important, then one must measure the full correlational structure. The latter can be extremely difficult, because it requires exponentially large amounts of data.

We should emphasize that even if correlations across multiple neurons turn out not to be important, simultaneous multineuron recording is still required for decoding the activity of populations of neurons. This is because it is essential to decode the true (correlated) responses, as these are the ones seen by the brain. Moreover, whether correlations are important for decoding or not, they may be important for other functions. For example, they may make postsynaptic neurons more likely to fire (11, 46–48).

Correlations are one of the major obstacles to cracking the neural code, as codes based on correlations can be tremendously complicated, whereas those based on independent responses are relatively simple. A method for assessing the role of correlations is thus a key step in understanding how stimuli are encoded in neuronal responses.

We acknowledge Stefano Panzeri for many useful discussions and Barry Richmond and Matt Wiener for helpful comments on the paper. This work was supported by National Eye Institute Grant R01 EY12978 and the Beckman Foundation (to S.N.) and National Institute of Mental Health Grant R01 MH62447 (to P.E.L.).

1. Adrian, E. D. & Zotterman, Y. (1926) *J. Physiol. (London)* **61,** 465–483.
2. Richmond, B. J., Optican, L. M., Podell, M. & Spitzer, H. (1987) *J. Neurophysiol.* **57,** 132–146.
3. Optican, L. M. & Richmond, B. J. (1987) *J. Neurophysiol.* **57,** 162–178.
4. de Ruyter van Stevinick, R. R. & Bialek, W. (1988) *Proc. R. Soc. London Ser. B* **234,** 379–414.
5. Victor, J. D. & Purpura, K. P. (1996) *J. Neurophysiol.* **76,** 1310–1326.
6. Brenner, N., Strong, S. P., Koberle, R., Bialek, W. & de Ruyter van Steveninck, R. R. (2000) *Neural Comput.* **12,** 1531–1552.
7. Milner, P. M. (1974) *Psychol. Rev.* **81,** 521–535.
8. von der Malsburg, C. (1981) *MPI Biophysical Chemistry: Internal Report 81-2*; reprinted in Domany, E., van Hemmen, J. L. & Schulten, K., eds. (1994) *Models of Neural Networks II* (Springer, Berlin).
9. von der Malsburg, C. (1985) *Ber. Bunsenges. Phys. Chem.* **89,** 703–710.
10. Gray, C. M. (1999) *Neuron* **24,** 31–47.
11. Reyes, A. (2001) *Annu. Rev. Neurosci.* **24,** 653–675.
12. Shadlen, M. N. & Movshon, J. A. (1999) *Neuron* **24,** 67–77.
13. Rodieck, R. W. (1967) *J. Neurophysiol.* **30,** 1043–1071.
14. Mastronarde, D. N. (1983) *J. Neurophysiol.* **49,** 303–324.
15. Mastronarde, D. N. (1983) *J. Neurophysiol.* **49,** 325–349.
16. DeVries, S. H. (1999) *J. Neurophysiol.* **81,** 908–920.
17. Gawne, T. J. & Richmond, B. J. (1993) *J. Neurosci.* **13,** 2758–2771.
18. Shannon, C. E. & Weaver, W. (1949) *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana).
19. Cover, T. M. & Thomas, J. A. (1991) *Elements of Information Theory* (Wiley, New York).
20. Nirenberg, S., Carcieri, S. M., Jacobs, A. L. & Latham, P. E. (2001) *Nature* **411,** 698–701.
21. Pola, G., Thiele, A., Hoffmann, K.-P. & Panzeri, S. (2003) *Network Comput. Neural Syst.* **14,** 35–60.
22. Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H. J. (1988) *Biol. Cybern.* **60,** 121–130.
23. Gray, C. M. & Singer, W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1698–1702.
24. Gray, C. M., Konig, P., Engel, A. K. & Singer, W. (1989) *Nature* **338,** 334–337.
25. Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. & Aertsen, A. (1995) *Nature* **373,** 515–518.
26. deCharms, R. C. & Merzenich, M. M. (1996) *Nature* **381,** 610–613.
27. Panzeri, S., Golledge, H. D. R., Zheng, F., Tovée, M. J. & Young, M. P. (2001) *Vis. Cognit.* **8,** 531–547.
28. Panzeri, S., Golledge, H. D. R., Zheng, F., Pola, G., Blanche, T. J., Tovée, M. J. & Young, M. P. (2002) *Neurocomputing* **44–46,** 579–584.
29. Liu, R. C., Tzonev, S., Rebrik, S. & Miller, K. D. (2001) *J. Neurophysiol.* **86,** 2789–2806.
30. Nirenberg, S. & Latham, P. E. (1998) *Curr. Opin. Neurobiol.* **8,** 488–493.
31. Oram, M. W., Foldiak, P., Perrett, D. I. & Sengpiel, F. (1998) *Trends Neurosci.* **21,** 259–265.
32. Panzeri, S., Pola, G., Petroni, F., Young, M. P. & Petersen, R. S. (2002) *BioSystems* **67,** 177–185.
33. Petersen, R. S., Panzeri, S. & Diamond, M. E. (2001) *Neuron* **32,** 503–514.
34. Dan, Y., Alonso, J. M., Usrey, W. M. & Reid, R. C. (1998) *Nat. Neurosci.* **1,** 501–507.
35. Oram, M. W., Hatsopoulos, N. G., Richmond, B. J. & Donoghue, J. P. (2001) *J. Neurophysiol.* **86,** 1700–1716.
36. Panzeri, S., Schultz, S. R., Treves, A. & Rolls, E. T. (1999) *Proc. R. Soc. London Ser. B* **266,** 1001–1012.
37. Panzeri, S., Treves, A., Schultz, S. & Rolls, E. T. (1999) *Neural Comput.* **11,** 1553–1577.
38. Panzeri, S. & Schultz, S. R. (2001) *Neural Comput.* **13,** 1311–1349.
39. Wu, S., Nakahara, H., Murata, N. & Amari, S. (2000) *Adv. Neural Inf. Process. Syst.* **11,** 167–173.
40. Wu, S., Nakahara, H. & Amari, S. (2001) *Neural Comput.* **13,** 775–797.
41. Heller, J., Hertz, J. A., Kjaer, T. W. & Richmond, B. J. (1995) *J. Comput. Neurosci.* **2,** 175–193.
42. Rieke, F., Bodnar, D. A. & Bialek, W. (1995) *Proc. R. Soc. London Ser. B* **262,** 259–265.
43. Strong, S. P., Koberle, R., de Ruyter van Stevinick, R. R. & Bialek, W. (1998) *Phys. Rev. Lett.* **80,** 197–200.
44. Reinagel, P. & Reid, R. C. (2000) *J. Neurosci.* **20,** 5392–5400.
45. Reich, D. S., Mechler, F. & Victor, J. D. (2001) *Science* **294,** 2566–2568.
46. Usrey, W. M., Reppas, J. B. & Reid, R. C. (1998) *Nature* **395,** 384–387.
47. Usrey, W. M., Alonso, J. M. & Reid, R. C. (2000) *J. Neurosci.* **20,** 5461–5467.
48. Nettleton, J. S. & Spain, W. J. (2000) *J. Neurophysiol.* **83,** 3310–3322.