

# The Role of Correlations in Direction and Contrast Coding in the Primary Visual Cortex

Fernando Montani,<sup>1</sup> Adam Kohn,<sup>2</sup> Matthew A. Smith,<sup>2</sup> and Simon R. Schultz<sup>1</sup>

<sup>1</sup>Department of Bioengineering, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom, and <sup>2</sup>Center for Neural Science, New York University, New York, New York 10003

The spiking activity of nearby cortical neurons is not independent. Numerous studies have explored the importance of this correlated responsivity for visual coding and perception, often by comparing the information conveyed by pairs of simultaneously recorded neurons with the sum of information provided by the respective individual cells. Pairwise responses typically provide slightly more information so that encoding is weakly synergistic. The simple comparison between pairwise and summed individual responses conflates several forms of correlation, however, making it impossible to judge the relative importance of synchronous spiking, basic tuning properties, and stimulus-independent and stimulus-dependent correlation. We have applied an information theoretic approach to this question, using the responses of pairs of neurons to drifting sinusoidal gratings of different directions and contrasts that have been recorded in the primary visual cortex of anesthetized macaque monkeys. Our approach allows us to break down the information provided by pairs of neurons into a number of components. This analysis reveals that, although synchrony is prevalent and informative, the additional information it provides frequently is offset by the redundancy arising from the similar tuning properties of the two cells. Thus coding is approximately independent with weak synergy or redundancy arising, depending on the similarity in tuning and the temporal precision of the analysis. We suggest that this would allow cortical circuits to enjoy the stability provided by having similarly tuned neurons without suffering the penalty of redundancy, because the associated information transmission deficit is compensated for by stimulus-dependent synchrony.

**Key words:** cerebral cortex; extracellular recording; information theory; neuronal ensembles; redundancy; striate cortex; synchronization; synchrony; synergy

## Introduction

A central question in neuroscience is understanding how information about the outside world is carried in neuronal spike trains. Information can be carried in spike rate (Werner and Mountcastle, 1965; Tolhurst, 1989), spike timing (Panzeri et al., 2001), spike correlations across neurons (Kreiter and Singer, 1992, 1996; De Charms and Merzenich, 1996; Gawne et al., 1996; Roelfsema et al., 1997), or a combination of these. Recently, a great deal of attention has been focused on correlated firing: that the probability of one cell spiking is related to whether other nearby cells fire (Mastrorarde, 1983; Ts'o and Gilbert, 1988; Engel et al., 1990; Gawne and Richmond, 1993; Zohary et al., 1994;

Kreiter and Singer, 1996; De Oliveira et al., 1997; Lebedev et al., 2000; Maldonado et al., 2000; Bair et al., 2001). Whether such correlations affect the coding of sensory information is still highly debated. It has been proposed that correlations might act as an extra channel for information, carrying messages about the outside world not carried by other aspects of spike trains, such as the overall firing rate (Richmond and Gawne, 1998; Abbott and Dayan, 1999; Gray, 1999; Panzeri et al., 1999). Others have proposed that correlation interferes with decoding the information represented by the firing rate of a population of neurons (Zohary et al., 1994; Shadlen and Newsome, 1998; Mazurek and Shadlen, 2002).

A quantitative answer to how correlation affects coding, independent of the how responses are decoded, can be provided by information theory. Unfortunately, information theoretic studies have provided disparate answers. For instance, retinal ganglion cells have been found to encode synergistically (Meister, 1996), approximately independently (Nirenberg et al., 2001), or redundantly (Puchalla et al., 2005). In the lateral geniculate nucleus Dan et al. (1998) reported synergistic effects. In cortex most studies have concluded that neurons provide approximately independent information, with some evidence for weak synergy in V1 (Reich et al., 2001; Golledge et al., 2003; Kayser and Konig, 2004), motor cortex (Oram et al., 2001; Averbeck and Lee, 2003), and somatosensory cortex (Petersen et al., 2001). Given that cor-

Received Aug. 8, 2006; revised Jan. 19, 2007; accepted Jan. 21, 2007.

This work was supported by a Howard Hughes Medical Institute Investigatorship and National Institutes of Health Grant EY2017 to J. Anthony Movshon and by grants from the Royal Society, Engineering and Physical Sciences Research Council (GR/S85467), and Gatsby Charitable Foundation to S.R.S. We thank Ines Samengo and Ilya Nemenman for useful discussions during the course of this research and Fabrizio Gabbiani, Ines Samengo, and Kit Longden for comments on a draft of this manuscript.

Correspondence should be addressed to Simon R. Schultz, Department of Bioengineering, Imperial College London, South Kensington, London SW7 2AZ, UK. E-mail: s.schultz@imperial.ac.uk.

A. Kohn's present address: Department of Neuroscience, Albert Einstein College of Medicine, 1410 Pelham Parkway South, Bronx, NY 10461.

M. A. Smith's present address: Center for the Neural Basis of Cognition, Carnegie Mellon University, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, PA 15213.

DOI:10.1523/JNEUROSCI.3417-06.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/272338-11\$15.00/0

tical responses are correlated strongly (Zohary et al., 1994; Bair et al., 2001; Reich et al., 2001) and that this correlation is stimulus-dependent (Kohn and Smith, 2005; Samonds and Bonds, 2005), it is surprising that the information provided by cortical neurons is nearly independent.

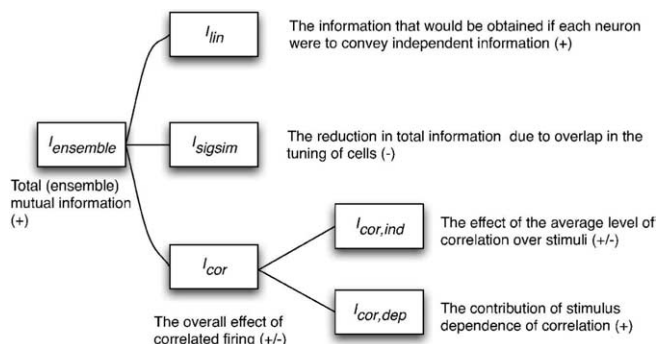
We have investigated the role of correlation in stimulus encoding in primary visual cortex, with two important extensions over previous studies. First, we use an exact information theoretic method to quantify the information conveyed by different coding mechanisms (Pola et al., 2003). Second, we apply this approach to responses evoked in primary visual cortex by stimuli that vary in two distinct ways: drift direction and contrast (Kohn and Smith, 2005). We confirm that coding is approximately independent despite the presence of substantial correlation, with weak synergy or redundancy arising for particular stimulus manipulations and time scales. We show that this independence comes about because of a balance between the strong synergy provided by stimulus-dependent correlation and redundancy arising from the similarity in tuning of the neurons.

### Materials and Methods

**Experimental procedures.** Recordings were made in 10 cynomolgus (*Macaca fascicularis*), one bonnet (*M. radiata*), and one pig-tailed (*M. nemestrina*) adult male monkeys. All experimental procedures were approved by the New York University Animal Welfare Committee. This data set has been described previously (Kohn and Smith, 2005), as have the experimental procedures used in the laboratory (Cavanaugh et al., 2002). Briefly, the animals were premedicated with atropine (0.05 mg/kg) and diazepam (1.5 mg/kg) and anesthetized initially with ketamine-HCl (10 mg/kg). Anesthesia during recording was maintained by intravenous infusion of sufentanil citrate (Sufenta; 4–8  $\mu\text{g} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ ). To minimize eye movements, we infused vecuronium bromide intravenously (Norcuron; 0.1 mg  $\cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ ). Vital signs (EEG, ECG, end-tidal  $\text{P}_{\text{CO}_2}$ , temperature, and lung pressure) were monitored continuously. The pupils were dilated with topical atropine, and the corneas were protected with gas-permeable contact lenses. Refraction was provided by supplementary lenses. Electrophysiological recordings were made via a seven-electrode array (Thomas Recording, Giessen, Germany). Spikes were detected by using a hardware discriminator and were digitized with a temporal resolution of 0.25 ms. Most pairs of cells (90.5%) were recorded on separate electrodes. Use of an anesthetized preparation for this experiment allowed many more trials to be collected than otherwise would be possible; this is particularly important for information theoretic analyses in which effective sampling is crucial. The use of sufentanil avoided some of the problems that are apparent with other anesthetics such as halothane and isoflurane, including changes in contrast sensitivity (Movshon et al., 2003) and oscillations (Imas et al., 2004).

**Visual stimuli.** Stimuli were luminance-modulated drifting sine-wave gratings presented at a frame rate of 100 Hz. We measured, in order, the direction, spatial and temporal frequency, and size tuning for drifting sine-wave gratings. After characterizing the stimulus preference for each cell independently, we measured correlation for responses evoked by stimuli of different directions and contrasts. The spatial and temporal frequency of these test stimuli were set between the preferred values of the two cells or at the value of the cell that responded less vigorously. Stimuli were presented between the receptive field (RF) centers of the cells and covered both RFs. All measurements were made by using stimuli presented in a circular aperture to the dominant eye of the less responsive cell. A gray field of average luminance surrounded the stimuli.

Direction and contrast experiments were performed in separate blocks of trials. Within most (~90%) of these experiments the presentation order of each direction or contrast was block randomized. In the direction experiments we presented full-contrast gratings drifting in five distinct directions, spanning the range between evoking a weak response and driving both cells strongly. In the contrast experiments we fixed the direction to that most effective at driving the pair of cells and presented the stimulus at four contrasts, typically 1.56, 6.25, 25, and 100%. For cells



**Figure 1.** Information component breakdown. The ensemble mutual information can be broken down into a linear component, the reduction of information attributable to the redundancy caused by overlap in tuning curves, and the contribution of correlated firing ( $I_{cor}$ ). The noise correlation term can be broken down additionally to separate out the effect of the average level of correlation over all stimuli and the stimulus dependence of correlation; the latter term captures any effects attributable to coding by explicit modulation of correlation/synchronization.

with low contrast sensitivity we adjusted the range of contrasts to extend from 12.5 to 100% in octave steps. Stimuli were presented for 2.56 s, with a 3 s interstimulus interval during which we presented an isoluminant gray screen. Each stimulus was presented 30–200 times. However, for the purposes of the information calculations (see below) these stimuli were divided into individual cycles of the drifting grating, and thus much greater numbers of experimental trials (480–3200; mean, 1300) were available for the purposes of information estimation.

**Response characterization.** We characterized the correlation by using the (shift predictor corrected) spike train cross-correlogram (CCG) [Perkel et al. (1967), and see supplemental material E, available at www.jneurosci.org] as described by Kohn and Smith (2005) as well as using information theory. For the purposes of the information calculations we used response bins with fixed time window lengths,  $T$  (which were integral fractions of the grating period  $\tau$ , i.e.,  $T = \tau$ ,  $T = \tau/2$ ,  $T = \tau/4$  etc., where  $\tau$  was the inverse of the stimulus temporal frequency used for the pair of cells ranging from 80 to 320 ms). Response vectors  $\mathbf{r}$  for each trial were calculated by counting spikes in these bins of length  $T$  ms; for codes comprising a pair of cells, the response vector had length two. The cardinality of the response vector was  $(n_{smax,1} + 1)(n_{smax,2} + 1)$ , where  $n_{smax,i}$  is the maximum number of spikes (over all stimuli, for any trial) that was observed to be fired by cell  $i$ . For each window length  $T$  we calculated the information quantities (as shown below) by using all 480–3200 trials (cycles). When multiple windows could be extracted from a single cycle (e.g., 4 windows at  $T = \tau/4$ ), the resulting information values from each window in the cycle were averaged together. Note that simple cells were cycle modulated, so the results should be taken to be average information estimates, which should not affect conclusions with regard to the relative effect of correlations.

**Theory.** In this section we describe how we calculated the Shannon (1948) information transmitted by neuronal population activity (and a number of related quantities). Shannon mutual information quantifies the extent to which the responses are dependent on the stimuli. It is a distance-like measure (the Kullback–Leibler divergence) between two probability distributions: the joint probability of responses and stimuli,  $P(\mathbf{r}, s)$ , and joint probability distribution that would apply if responses did not depend at all on stimuli,  $P(\mathbf{r})P(s)$ . This dependence between responses and stimuli may take a number of forms: firing rate dependence, pairwise correlation dependence, etc. Information component breakdown methods (Fig. 1) allow us to quantify the effect of such dependencies on the mutual information and thus assess the different ways in which the correlations contribute to the neural code. If synergistic or redundant interactions between cells are observed, the information components may reveal the mechanisms from which they arise.

The technical approach we adopt is a modification of a previously developed technique (Panzeri et al., 1999; Panzeri and Schultz, 2001; Schultz and Panzeri, 2001) in which a Taylor series expansion was used to

break the Shannon (1948) information into its components. Pola et al. (2003) generalized this approach by substituting correlation functions with probability functions, resulting in an exact information component breakdown. One disadvantage to the Pola approach is that the correlation components cannot be written explicitly in terms of entropies of probability distributions. Although we do not see this as giving rise to problems of interpretation (but see Schneidman et al., 2003), it does mean that advanced methods for entropy estimation cannot be used. Thus we have rewritten the correlational components as approximate entropies (i.e., an entropy plus additional terms that are very small and relatively immune to bias), allowing us to use a recently developed technique for minimizing the bias associated with estimating entropy from a limited number of experimental samples (Nemenman et al., 2004). By using a Bayesian prior to generate a nearly uniform distribution of entropies, we thus correct for sample size-dependent bias at its source and avoid potential artifacts that may occur when sampling is insufficient, as is typically the case in neurophysiological experiments.

We consider a time period of duration  $T$  unit time bins in which the activity of a given pair of cells is observed. The neural population response will be denoted by  $\mathbf{r}$  (drawn from a response space  $R$ ), and a sensory stimulus from a given stimulus set  $S$  will be denoted by  $s$ . The mutual information transmitted by the population response about the whole set of stimuli (Shannon, 1948; Cover and Thomas, 1991) is written as follows:

$$I(R;S) = H(R) - H(R|S), \quad (1)$$

where  $H(R)$  and  $H(R|S)$  are the total response entropy and the noise entropy, respectively. They are defined as follows:

$$H(R) = - \sum_{\mathbf{r} \in R} P(\mathbf{r}) \log_2 P(\mathbf{r}), \quad (2)$$

and

$$H(R|S) = - \sum_{s \in S} P(s) \sum_{\mathbf{r} \in R} P(\mathbf{r}|s) \log_2 P(\mathbf{r}|s), \quad (3)$$

where  $P(\mathbf{r}|s)$  is the probability of observing a given ensemble response vector  $\mathbf{r}$  conditional on the occurrence of stimulus  $s$ , and  $P(\mathbf{r})$  is the average of  $P(\mathbf{r}|s)$  over all stimuli.

To understand the meaning of correlation, one must define the probability of getting independent population responses as the following:

$$P_{\text{ind}}(\mathbf{r}|s) = \prod_{c=1}^C \prod_{t=1}^T P(r_{ct}|s), \quad (4)$$

and

$$P_{\text{ind}}(\mathbf{r}) = \langle P_{\text{ind}}(\mathbf{r}|s) \rangle_s, \quad (5)$$

where  $c$  is the label for each cell (up to  $C = 2$  cells in the pair for the analysis presented here, although the formalism is quite general in this respect),  $t$  indexes the time bin up to a maximum value of  $L$  corresponding to duration  $T$ , and the stimulus average is the following:

$$\langle x \rangle_s = \sum_{s=1}^S P(s) x. \quad (6)$$

The presence of noise correlation (correlation in the response variability for a fixed stimulus) or signal correlation (correlation in the tuning of, or signal conveyed by response variables) is indicated by  $P_{\text{ind}}(\mathbf{r}|s) \neq P(\mathbf{r}|s)$  and  $P(\mathbf{r}) \neq P_{\text{ind}}(\mathbf{r})$ , respectively. Moreover,  $P(s|\mathbf{r})$  is the true distribution of stimuli given responses, and  $P_{\text{ind}}(s|\mathbf{r})$  is the distribution one would derive in the absence of knowledge of correlations. In practice, we obtained  $P_{\text{ind}}(\mathbf{r}|s)$  numerically by multiplying the marginal probability distributions, as is apparent from Equation 4.

The information component breakdown method allows us to write the total mutual information into a sum of components that are related to

the different decoding mechanism (Panzeri and Schultz, 2001; Pola et al., 2003):

$$I(R;S) = I_{\text{lin}} + I_{\text{sig}} + I_{\text{cor}} \quad (7)$$

The first term of the information breakdown,  $I_{\text{lin}}$ , gives the total amount of information that would be conveyed if all of the cells were independent:

$$I_{\text{lin}} = \sum_{c=1}^C \sum_{t=1}^T [H(R_{ct}) - H(R_{ct}|S)] \quad (8)$$

where

$$H(R_{ct}|S) = - \sum_{s \in S} P(s) \sum_{r_{ct}} P(r_{ct}|s) \log_2 P(r_{ct}|s), \quad (9)$$

and  $H(R_{ct})$  is the averaged value across all of the stimuli.

The signal similarity term  $I_{\text{sig}}$  quantifies the information loss arising from redundancy because of an overlap in the tuning curves pertaining to response (cell, time) bins  $r_{ct}$ :

$$I_{\text{sig}} = H_{\text{ind}}(R) - \sum_{c=1}^C \sum_{t=1}^T H(R_{ct}), \quad (10)$$

where

$$H_{\text{ind}}(R) = - \sum_{\mathbf{r}} P_{\text{ind}}(\mathbf{r}) \log_2 P_{\text{ind}}(\mathbf{r}), \quad (11)$$

The third term,  $I_{\text{cor}}$ , quantifies the total amount of information attributable to the correlated activity on the overall neural coding:

$$I_{\text{cor}} = I(R;S) - H_{\text{ind}}(R) + \sum_{c=1}^C \sum_{t=1}^T H(R_{ct}), \quad (12)$$

So far, it is apparent that all of the terms can be written in terms of entropies of particular distributions. However, this correlation term can be resolved, in addition, into two components, a stimulus-independent component  $I_{\text{corind}}$  and a stimulus-dependent component  $I_{\text{cordep}}$ . This last term in the Pola et al. (2003) formalism is calculated as the following:

$$I_{\text{cordep}} = I(R;S) + \chi + \sum_c H(R_c|S), \quad (13)$$

where

$$\chi = - \sum_{\mathbf{r}} P(\mathbf{r}) \log_2 P_{\text{ind}}(\mathbf{r}). \quad (14)$$

That is, the stimulus-dependent correlation component is not made up entirely of entropies, and thus advanced entropy estimation techniques such as that by I. Nemenman, F. Shafee, and W. Bialek (NSB) (Nemenman et al., 2002, 2004) cannot be applied. Moreover,  $\chi$  can be biased substantially, which renders its calculation by previously available procedures difficult for all except very low-dimensional problems.

Formally speaking, the stimulus-dependent correlation component can be defined as the Kullback–Leibler divergence between  $P(s|\mathbf{r})$  and  $P_{\text{ind}}(s|\mathbf{r})$  (Nirenberg et al., 2001; Latham and Nirenberg, 2005), conceptually characterized by Nirenberg and colleagues as the effect of correlations on the decoding of stimuli as follows:

$$I_{\text{cordep}} = D(P(s|\mathbf{r})||P_{\text{ind}}(s|\mathbf{r})) = \sum_{\mathbf{r}} P(\mathbf{r}) \sum_s P(s|\mathbf{r}) \log_2 \frac{P(s|\mathbf{r})}{P_{\text{ind}}(s|\mathbf{r})}, \quad (15)$$

In the following we rewrite the correlation stimulus-dependent component from “first principles” by using Kullback–Leibler for the case in which only pairwise correlations between neurons are considered. We

note that this formalism easily can be extended to a larger number of cells as follows:

$$I_{\text{cordep}} = D(P(s|r_1r_2)||P_{\text{ind}}(s|r_1r_2)) \equiv \sum_{r_1r_2} \sum_s P(s) P(r_1r_2|s) \log_2 \frac{P(r_1r_2|s)}{P_{\text{ind}}(r_1r_2|s)} - \sum_{r_1r_2} P(r_1r_2) \log_2 \frac{P(r_1r_2)}{P_{\text{ind}}(r_1r_2)} \quad (16)$$

defining

$$\beta = \frac{\langle P(r_1|s) \rangle_s \langle P(r_2|s) \rangle_s}{\langle P(r_1|s) \rangle_s \langle P(r_2|s) \rangle_s}; \quad (17)$$

The correlation stimulus-dependent component can be rewritten as follows:

$$I_{\text{cordep}} = \sum_{r_1r_2} \sum_s P(s) P(r_1r_2|s) \log_2 \frac{P(r_1r_2|s)}{P_{\text{ind}}(r_1r_2|s)} - \sum_{r_1r_2} P(r_1r_2) \log_2 \frac{P(r_1r_2)}{\langle P(r_1|s) \rangle_s \langle P(r_2|s) \rangle_s} - \Delta, \quad (18)$$

where

$$\Delta = \sum_{r_1r_2} P(r_1r_2) \log_2 \beta. \quad (19)$$

The stimulus-dependent correlation component thus has been expressed in terms of entropy quantities with the addition of a single term  $\Delta$ ; the NSB estimation method can be applied to these quantities with the exception of  $\Delta$ . The remaining non-entropy term ( $\Delta$ ) is essentially the stimulus-averaged response multiplied by a logarithmic function of the coefficient  $\beta$  summed over the whole space of responses. We are summing therefore in the space of responses quantities that have been averaged over the stimuli, and it is reasonable to expect a very low bias contribution from this term. Moreover, because  $\beta$  is the product of the stimulus-averaged response from each cell divided by the stimulus unconditional independent responses, it is reasonable to expect that it will always be close to one, and therefore  $\Delta$  will be close to zero (supplemental material A, available at [www.jneurosci.org](http://www.jneurosci.org)). In practice, we included  $\Delta$  in our calculations (without bias correction); the magnitude of  $\Delta$  was always observed to be small.

Having expressed all of the quantities necessary to perform the information component analysis in terms of entropies of particular (in some cases conditional) distributions, we computed these entropies by inserting the respective distributions into the NSB entropy estimation routine (supplemental material B, available at [www.jneurosci.org](http://www.jneurosci.org)). We found that, for our data set, this procedure resulted in substantially better estimation performance than we were able to achieve with techniques that we have used previously (Panzeri and Treves, 1996). In addition, the NSB approach provides a convenient error estimate for the entropies, which can be propagated into error estimates for the information quantities (supplemental material C, available at [www.jneurosci.org](http://www.jneurosci.org)).

We have expressed all information quantities in total information terms, measured in bits, as opposed to rates measured in bits per second. In the context of the present study we believe that this leads to a clearer interpretation; for a discussion of the relationship to information rate quantities, see supplemental material D (available at [www.jneurosci.org](http://www.jneurosci.org)).

## Results

We analyzed the activity of 147 pairs of single neurons in the primary visual cortex of anesthetized, paralyzed macaque monkeys (Kohn and Smith, 2005). The neurons in each pair were typically complex cells separated by  $<500 \mu\text{m}$  that had a mean receptive field overlap of 75% and similar tuning properties: a mean difference of  $37^\circ$  in direction preference, 0.37 octaves in spatial frequency preference, and 0.36 octaves in temporal fre-

quency preference. These neuronal pairs previously have been shown to manifest stimulus-dependent synchrony (Kohn and Smith, 2005). In this paper we present an information theoretic analysis of the effect of this synchrony on neural coding of direction and contrast, making use of a modified version of the method of information components (Panzeri and Schultz, 2001; Pola et al., 2003). We report here on a subset of 102 pairs (for the orientation analysis) and 71 pairs (for the contrast analysis) of the original Kohn and Smith (2005) data set, chosen to satisfy convergence criteria for the entropy estimation procedure that we used (Nemenman et al., 2004).

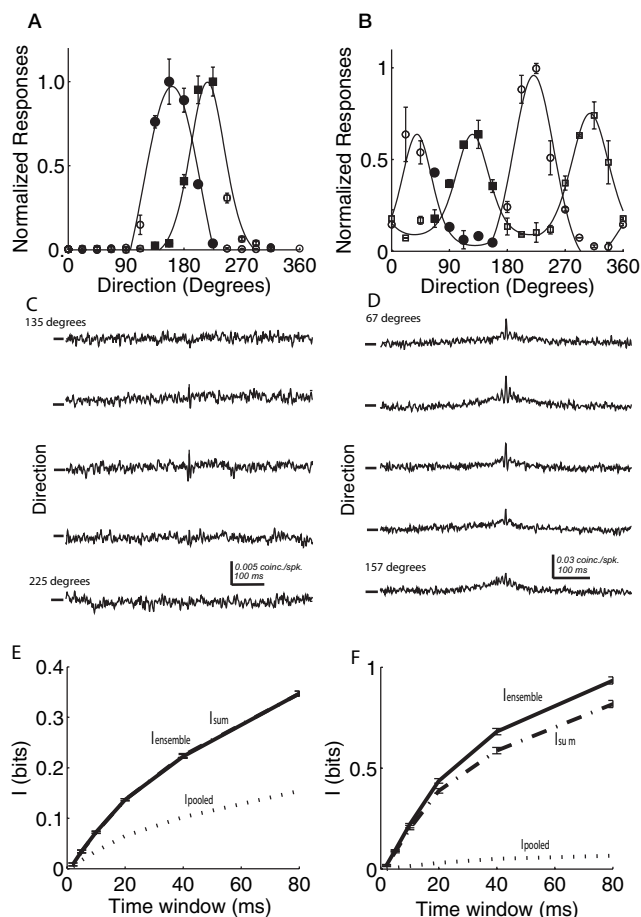
### Direction coding: information adds linearly across cells

We examined the population coding of stimulus direction in primary visual cortex by quantifying simultaneously recorded pairwise responses as the number of spikes fired by each cell in the pair within a time window  $T$  (a “two letter word” response). We then calculated the total mutual information conveyed by the ensemble response about which stimulus direction gave rise to the response (see Materials and Methods). We designated this  $I_{\text{ensemble}}$  (Fig. 1) and compared it with two other information measures: the sum of the mutual information calculated from the responses of the individual neurons (i.e., the sum of two pieces of information calculated from two “one letter words,”  $I_{\text{sum}}$ ) and the mutual information calculated from a reduced code in which the spikes on each trial are pooled across cells [i.e., the information from a single one letter word,  $I_{\text{pooled}}$  (Reich et al., 2001)]. If the responses of the neurons are uncorrelated completely, then  $I_{\text{ensemble}}$  should be equal to  $I_{\text{sum}}$ ; in the presence of correlations  $I_{\text{ensemble}}$  might exceed  $I_{\text{sum}}$ , which we would refer to as a synergistic interaction, or might be less than  $I_{\text{sum}}$ , which we would refer to as a redundant interaction. This commonly used definition of redundancy between a pair of information channels (cells) measures the extent to which they carry common information; it is not the same as the redundancy within a single information channel introduced by Shannon (1948), which is defined in terms of information capacity.

Figure 2A shows a typical pair of cells (which we label pair A) with partially overlapping tuning curves (their direction preferences were  $49^\circ$  apart). We presented five stimuli (Fig. 2A, filled symbols) that drove each cell through a wide range of firing rates. The result, as can be seen in Figure 2C, is a substantial modulation in the height of the central peak of the CCG of the spike trains of the neurons; at a particular orientation that drives both neurons moderately well, a fraction of spikes fired by each neuron tends to be synchronized with millisecond temporal precision (Kohn and Smith, 2005). For stimuli that do not drive both cells well, the extent of synchronization falls off sharply. This stimulus dependence of synchrony arises despite correcting for the basic rate dependence of the CCG (Kohn and Smith, 2005); the origin of the stimulus dependence in the CCG is discussed in supplemental material E (available at [www.jneurosci.org](http://www.jneurosci.org)).

The synchronized firing illustrated above is an example of correlation. The timing of spikes in the two neurons is not independent. Does the substantial synchronization apparent in this example lead to a synergistic or redundant encoding of stimulus direction? In Figure 2E it can be seen that the answer for this pair is neither. In this case there appears to be a balance of contributions such that the overall effect of the correlation is neither synergistic nor redundant. The ensemble information ( $I_{\text{ensemble}}$ ) is exactly equal to the sum of the single cell information values ( $I_{\text{sum}}$ ) for a wide range of time windows.

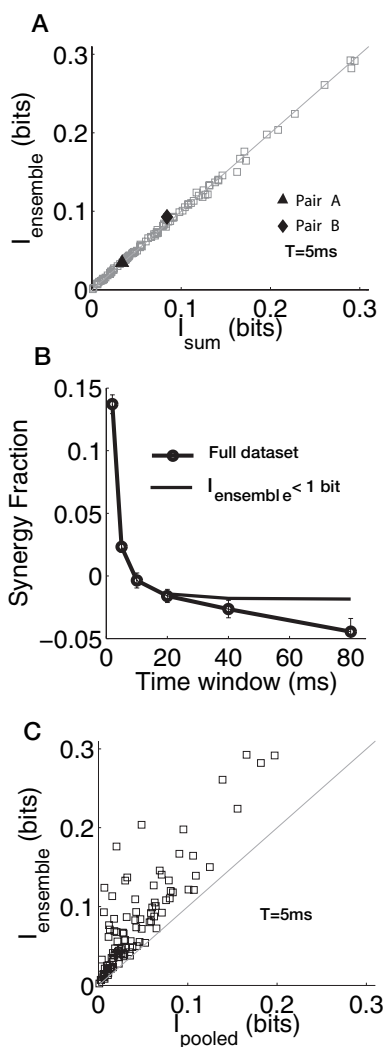
A second example pair with more widely separated direction



**Figure 2.** Direction coding by two pairs of neurons. *A, B*, Direction tuning curves for two pairs of neurons. Both of the first pair are direction-selective and the second only slightly directionally biased (fit with von Mises functions). Error bars indicate SEM. *C, D*, CCGs for each of the pairs above show that fine time scale synchronization is induced for stimulus directions that drive both cells relatively well. The cross-correlation was measured for the five stimuli indicated by the filled symbols in *A* and *B*. *E, F*, Information analysis. The total information available from each pair of neurons ( $I_{ensemble}$ ) is compared for different integration time windows,  $T$ , to the sum of the information values obtained from each constituent neuron alone ( $I_{sum}$ ) and to the information from a pooled code in which the identity of the cell firing each action potential is ignored ( $I_{pooled}$ ).

tuning curves (difference in direction preference of  $87^\circ$ ) is shown in Figure 2*B* (which we label pair B). The height of the central peak of the CCG is again sensitive to stimulus direction (Fig. 2*D*), but in this case the resulting interaction is synergistic; the ensemble mutual information is larger than the sum regardless of the time window over which spikes are counted (Fig. 2*F*).

The picture over the entire population is one of a small amount of synergy; with a time window of 5 ms the ensemble code does on average 2.4% better than the sum of single cell contributions (Fig. 3*A, B*). To compute the degree of synergy in the population for a range of time windows  $T$ , we calculated a synergy index (which we call the synergy fraction) as  $(I_{ensemble} - I_{sum})/I_{ensemble}$ ; values  $> 0$  indicate synergistic coding, and values  $< 0$  indicate redundancy. The synergistic effect of ensemble coding is more substantial at shorter time windows, where the fine temporal precision at which the spikes may synchronize has a significant effect. The maximum effect we observed was 13.7% at 2 ms (for technical reasons related to convergence of the numerical integral involved in the entropy computation, this was the smallest time window we were able to use), tailing off sharply



**Figure 3.** Information adds relatively linearly across cells. *A*, The total (ensemble) information available for each pair is plotted against the sum of the information available individually from each cell constituting the pair. The time window used for counting spikes was 5 ms. Pairs marked A and B correspond to the pairs illustrated in Figure 2, *A* and *B*. *B*, The synergy fraction (defined as  $1 - I_{sum}/I_{ensemble}$ ) and indicating the fraction of information beyond that expected if the cells are independent) is plotted against the time window used for counting spikes. The lines show the average  $\pm$  SEM over the entire data set and over only those pairs showing ensemble information  $< 1$  bit (this latter curve excluding potential artifactual redundancy effects). *C*, Pooling spikes across cells destroys a substantial amount of information in almost all cases, indicating that the pattern of spikes across cells is informative.

at  $\sim 10$  ms, a point at which the synergy fraction is close to zero (Fig. 3*B*). Thus being very precise about spike timing leads to synergy rather than information independence (Schneidman et al., 2003). The short time scale over which synergy is present agrees well with the time scale of synchrony in the population; the mean full width at half-maximum of the CCG peaks was 9 ms. At longer time windows ( $T \geq 20$  ms) the information became comparable in size to the stimulus entropy ( $\log_2, 5$  bits), and thus ceiling effects may reduce the raw information values somewhat. We corrected for this by removing pairs from the population that presented a higher information amount than 1 bit at a given time window, an arbitrary but conservative threshold. It is apparent that in the absence of ceiling effects the trend for long time windows is for the synergy fraction to asymptote to a slightly negative value, indicating that the mutual information adds across the neurons as if they were approximately independent.

In contrast to  $I_{\text{sum}}$ , the information available from the pooled code ( $I_{\text{pooled}}$ ) is curtailed severely; destroying the identity of which cell fired a spike substantially reduces the information transmitted [in agreement with Reich et al. (2001)]. This was apparent in both example pairs A and B (Fig. 2*E, F*) and is true for every pair that was analyzed (Fig. 3*C*). The mean amount of information lost that was caused by destroying cell identity was 46.33% of the ensemble information at  $T = 5$  ms and 46% at  $T = 40$  ms (Reich et al., 2001).

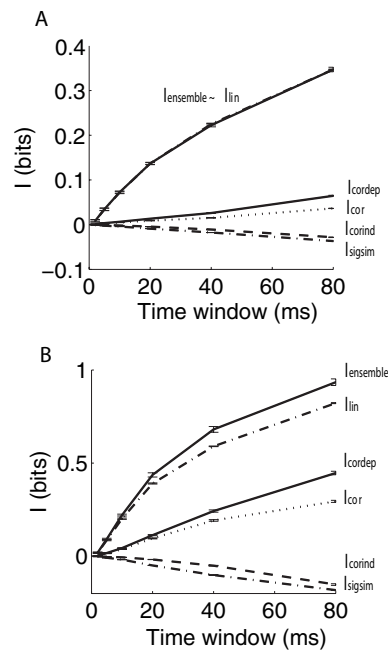
### The role of correlations in direction coding: a balance of redundant and synergistic effects

In the previous section we found that correlations between neurons led to a weakly synergistic code for direction on fine time scales. In a sense, this is surprising because the pairs often displayed substantial synchronization for which the strength was strongly dependent on the direction of the stimulus (Fig. 2*C, D*) (Kohn and Smith, 2005), suggesting that this synchrony could convey substantial information. On the other hand, the synchrony arises from common input to a pair of cells and was prevalent in our data set precisely because the neurons had similar receptive field properties. The similarity in tuning properties means that the firing rate provides redundant information and raises the issue of how this information compares with that provided by correlated firing.

To examine the relative information conveyed by different aspects of the response, we separated the total mutual information into components reflecting the contributions of individual coding mechanisms (Fig. 1). The linear component  $I_{\text{lin}}$  represents the information that would be obtained if each neuron were to convey independent information. This is equal to  $I_{\text{sum}}$ , which we calculated in the previous section by a different approach.  $I_{\text{sigstim}}$  represents the reduction in total information conveyed because of one source of redundancy: the overlap in the tuning of the cells. These first two components depend only on the firing rates of the individual neurons, rather than the correlated or synchronized firing between neurons. The  $I_{\text{cor}}$  term captures the consequences of this correlated firing and can be separated additionally into stimulus-dependent ( $I_{\text{cordep}}$ ) and stimulus-independent ( $I_{\text{corind}}$ ) parts. The effect of  $I_{\text{cordep}}$ , if non-zero, is always positive, but the effect of  $I_{\text{corind}}$  can be either positive or negative (synergistic or redundant), depending on the sign and magnitude of the overlap in tuning curves (Panzeri et al., 1999). Breaking the information into its respective components allows us to identify how synergy or redundancy arises.

Figure 4*A* shows the breakdown into components of the mutual information for pair A. In this case the ensemble information is exactly equal to the linear component, indicating that there is (overall) no information provided by the correlation (consistent with the analysis of Fig. 2*E*). This is true despite the presence of stimulus-dependent synchrony in this pair (Fig. 2*C*). The reason for this lies in redundant contributions from the stimulus-independent correlation ( $I_{\text{corind}}$ ) and the overlap in the tuning of the two cells,  $I_{\text{sigstim}}$ , which offset exactly the information provided by the stimulus-dependent synchrony ( $I_{\text{cordep}}$ ). Pair B, which showed a stronger synchronization effect, also showed a stronger contribution of the correlational components of the information. In this case the stimulus-dependent correlation component  $I_{\text{cordep}}$  was sufficiently large to outweigh redundant effects attributable to the average level of correlation ( $I_{\text{corind}}$ ) and the similarity in tuning ( $I_{\text{sigstim}}$ ), leading to an overall synergistic effect of correlation (Fig. 4*B*).

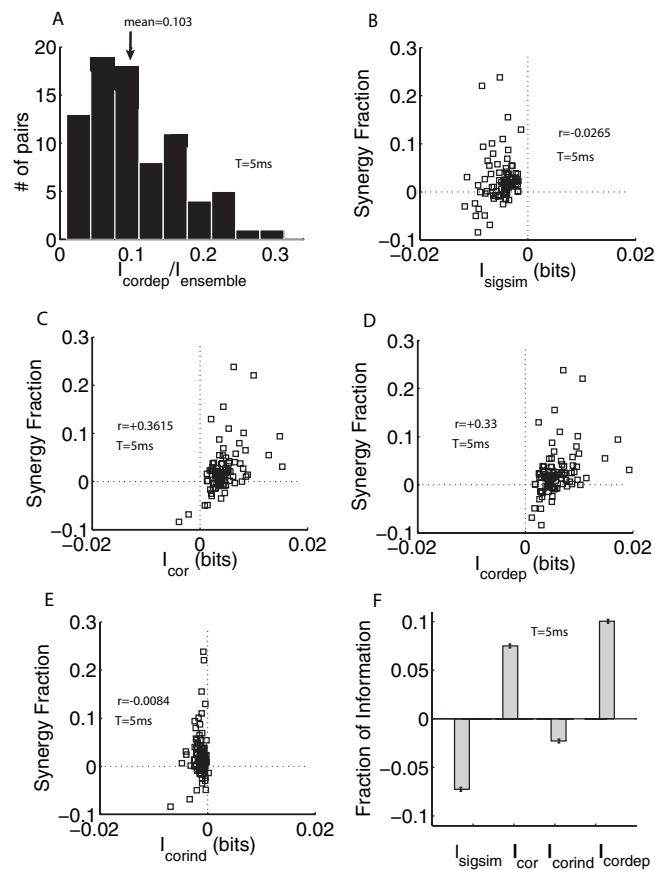
For pair A (a fairly typical pair of cells), synchronization af-



**Figure 4.** Information component analysis for the example pairs, showing how the balance between a synergistic contribution attributable to orientation-dependent synchronization ( $I_{\text{cordep}}$ ) and redundant contributions from tuning overlap ( $I_{\text{sigstim}}$ ) and the average level of correlation ( $I_{\text{corind}}$ ) gives rise to the overall level of synergy or redundancy. The information components  $I_{\text{lin}}$ ,  $I_{\text{sigstim}}$ , and  $I_{\text{cor}}$  combine to give the total information,  $I_{\text{ensemble}}$ . **A, B** correspond to the same pairs of cells described in Figure 2, **A** and **B**.

ected information content, but the stimulus dependence of the synchronization served to balance the redundancy precisely because of overlapping tuning. The value of  $I_{\text{cordep}}$  [called  $\Delta I$  by Nirenberg et al. (2001), derived by using a different approach] has been argued to indicate the importance of correlations for decoding (Latham and Nirenberg, 2005):  $I_{\text{cordep}}$  can be written as the Kullback–Leibler divergence between  $P(s|\mathbf{r})$ , the probability that a stimulus  $s$  elicited ensemble response  $\mathbf{r}$ , and  $P_{\text{ind}}(s|\mathbf{r})$ , such a probability distribution constructed without knowledge of the correlations. If correlations are not important for decoding which stimulus gave rise to the response, then these probability distributions should be identical, and  $I_{\text{cordep}}$  should be zero; as correlations increase in importance for decoding,  $I_{\text{cordep}}$  also should increase. For pair A the value of  $I_{\text{cordep}}$  was relative small (10% of the ensemble information at  $T = 5$  ms). For pair B the explicit information content in the stimulus dependence of correlations outweighed redundant effects and led to a more substantive value for  $I_{\text{cordep}}$  (18.5% at  $T = 5$  ms). For time windows of 5 ms the mean value of  $I_{\text{cordep}}$  over the population was 10.3% (Fig. 5*A*). For longer time windows ( $T = 40$  ms) the mean value of  $I_{\text{cordep}}$  was also  $\sim 10\%$ .

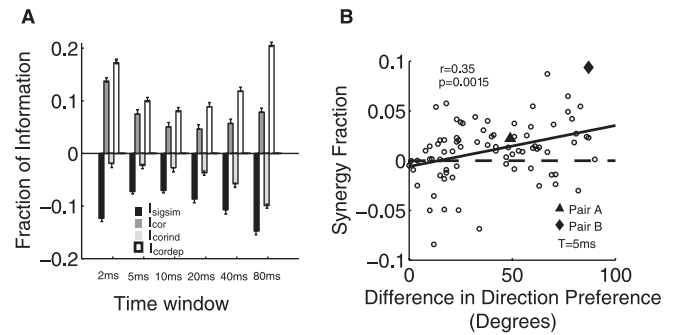
Which information components were responsible for the small amount of synergy that was observed? We examined this point by comparing the fraction of synergistic information  $(I_{\text{ensemble}} - I_{\text{sum}})/I_{\text{ensemble}}$  with the information components on a pair-by-pair basis. For  $T = 5$  ms it can be seen that the redundancies added by the signal similarity term  $I_{\text{sigstim}}$  (Fig. 5*B*) are canceled mostly by synergistic contributions of the total correlation component  $I_{\text{cor}}$  (Fig. 5*C*).  $I_{\text{cor}}$  is related significantly to the synergy fraction (correlation coefficient,  $r = +0.36$ ; significant to  $p = 0.0005$  by transforming the correlation to generate the  $t$  statistic) and contributes mostly synergistically (Fig. 5*C*). The component breakdown of  $I_{\text{cor}}$  reveals that  $I_{\text{cordep}}$  (Fig. 5*D*) is most



**Figure 5.** Information components across the entire population of 102 pairs (direction coding). *A*, Histogram of the relative contribution of  $I_{\text{cordep}}$  to the total information for a time window of 5 ms. *B*,  $I_{\text{sigsim}}$ , reflecting the redundant contribution attributable to the overlap in tuning curves, is generally non-zero but is not correlated strongly with the overall amount of synergy or redundancy. *C*, Correlational effects captured by  $I_{\text{cor}}$  are related more strongly to the extent of synergy in the coding. *D*, The stimulus-dependent correlational component is related strongly to the extent of synergistic coding, whereas (*E*) the average level of correlation is unrelated (*E*). *F*, A bar chart capturing the average value of each information component (expressed as a fraction of the total information,  $I_{\text{ensemble}}$ ) over the entire data set. Error bars indicate SEM ( $n = 102$ ).

predictive of the extent to which the pairs of neurons interact synergistically ( $r = +0.33$ ;  $p = 0.001$ ); in contrast,  $I_{\text{corind}}$  is unrelated to the synergy fraction ( $r = -0.008$ ;  $p = 0.94$ ) (Fig. 5*E*). The average value of each of the components over the entire population of pairs is shown in Figure 5*F* for a time window of 5 ms and in Figure 6*A* for a range of time windows.

In summary, it is the balance between  $I_{\text{cordep}}$  on the one hand, and  $I_{\text{sigsim}}$  and  $I_{\text{corind}}$  on the other, that mainly determines the degree of synergy. For time windows lower than 10 ms, the preservation of temporally precise spike synchrony results in correlational information that is a relatively large fraction of the total information  $I_{\text{ensemble}}$  (Fig. 6*A*). At time windows longer than 10 ms the synergy contribution from  $I_{\text{cordep}}$  is curtailed significantly and effectively canceled by the signal similarity term  $I_{\text{sigsim}}$  and also by  $I_{\text{corind}}$  which provides a greater redundant contribution as the time window increases (Fig. 6*A*). This is in agreement with the study of Kohn and Smith (2005), in which an orientation-sensitive component of the response correlation was found to dominate for short time scales and an orientation-insensitive component was found to dominate for longer time scales.



**Figure 6.** *A*, The relative effects of the information component are time scale dependent. Bar charts show the average fraction of the information accounted for by the different components for time windows ranging from 2 to 80 ms; whereas the redundant contribution of  $I_{\text{corind}}$  grows with time window, the effects of the other components follow a U or inverted U curve. Error bars indicate SEM. *B*, Wide separation in direction preference leads to synergy. Time windows,  $T = 5$  ms. Direction coding is redundant for pairs of neurons with similar tuning (negative synergy fraction).

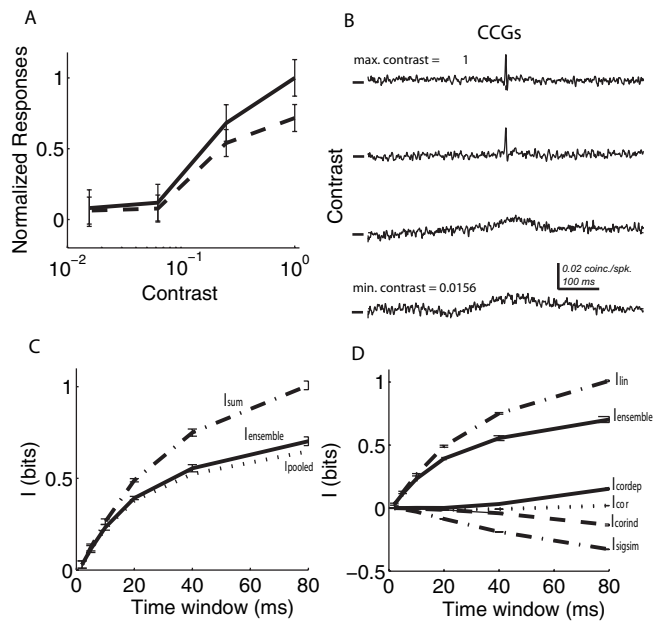
### Greater separation in direction preference leads to synergy

One important difference between the two examples is that for pair A the tuning preferences are closer together than for pair B, although pair B showed a stronger synergistic relationship than did pair A. As shown in Figure 6*B* for a time window of 5 ms, this is indicative of a general relationship; pairs with a greater separation in their preferred direction were more likely to interact synergistically, and those with preferred directions close together were more likely to carry redundant information. Linear summation of information across cells, when the synergy fraction is zero, occurred for neurons for which the preferred directions were  $\sim 14^\circ$  apart (linear fit;  $r = 0.35$ ;  $p = 0.0015$ ) (Fig. 6*B*) for a time window of 5 ms. Neurons with widely separated tuning curves thus carry complementary information, even when correlations are taken into account, whereas neurons with similar direction tuning curves tend to carry redundant information.

The finding that synergy is strongest for pairs with dissimilar tuning may appear surprising because correlated firing, and sharp synchrony in particular, are strongest between nearby neurons that have similar tuning (Nelson et al., 1992; Lee et al., 1998; De Angelis et al., 1999; Nowak et al., 1999; Bair et al., 2001; Kohn and Smith, 2005). Our analysis reveals, however, that synergy is strongest between pairs of neurons for which correlation would be expected to be relatively weak. This is because, although nearby neurons are particularly strongly correlated, this correlation arises from strong common input that presumably also gives rise to strongly overlapping tuning curves. The redundancy that arises from having similar tuning outweighs the information provided by correlation. For pairs with different preferences even weak correlation can lead to synergistic coding because the tuning of the cells is less redundant. However, correlations do not necessarily lead to synergy. Weak correlations can lead to synergy or redundancy, depending on the similarity in tuning and relative magnitudes of  $I_{\text{cordep}}$  and  $I_{\text{corind}}$ . Attempting to relate the strength of stimulus-dependent synchrony to synergistic coding is thus perilous. When similarity in tuning and stimulus-independent correlation are taken into account, it is not necessarily the case that stronger synchrony leads to more synergistic coding.

### Contrast coding is redundant

In the previous sections we have used information theory to study the role of correlations in the coding of stimulus direction in primary visual cortex. However, correlations also are modulated

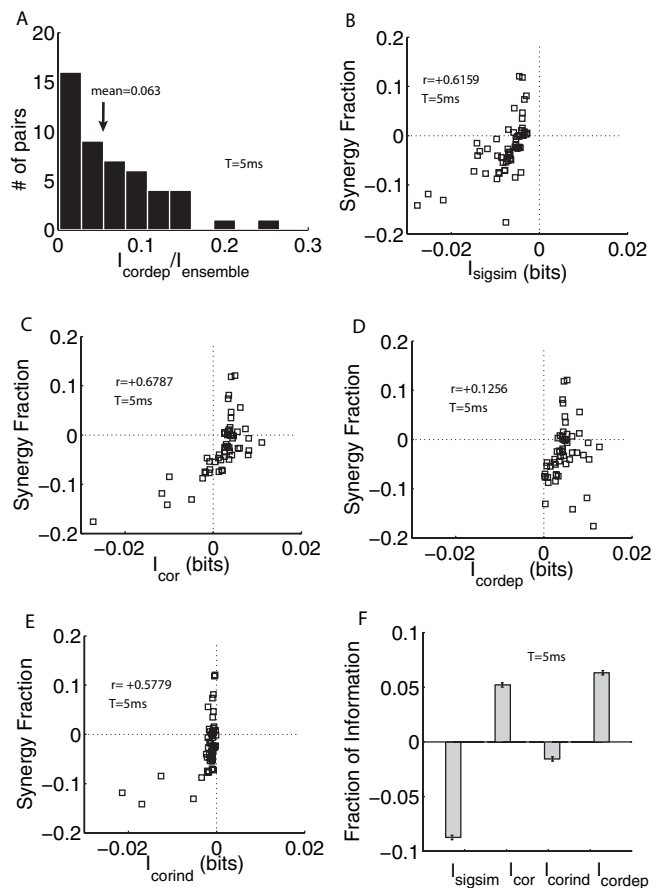


**Figure 7.** Contrast coding. *A*, Contrast tuning for a typical pair of cells (normalized to the maximal response of one of the cells). Error bars indicate SEM. *B*, CCGs show a decrease in the precision of spiking synchrony as contrast decreases. *C*, The ensemble information is lower than the sum of that available from each individual cell; contrast is coded redundantly by this pair of neurons. *D*, Information components. The overall effect of  $I_{cor}$  is minimal for this example; the redundant effect can be seen to stem from  $I_{sigstim}$  reflecting the strong overlap in contrast tuning.

by stimulus contrast (Kohn and Smith, 2005); whereas direction determines the height of the central peak of the CCG, the width of the central peak is primarily sensitive to stimulus contrast. This can be seen in the example in Figure 7, a pair of neurons with relatively similar contrast-tuning curves (Fig. 7*A*), which show a broadening of the width of the CCG as contrast is reduced (Fig. 7*B*). The different effect of altering stimulus direction and contrast on correlation provides an important test of whether there is a general role for correlation in stimulus coding or whether the importance of correlation depends on the particular stimulus attribute in question.

In the example of Figure 7, the ensemble information about contrast is substantially less than the sum of the information available from each cell independently, regardless of the window of analysis (Fig. 7*C*). Thus the coding of contrast is redundant. This can be explained by the fact that the contrast-tuning functions of the neurons are similar, the values of stimulus-dependent correlation component are smaller for contrast, and there is a significant level of stimulus-independent correlation (e.g., substantial synchrony that is present at all contrast levels). Figure 7*D* shows that the major contributor in the coding of contrast is the large redundant  $I_{sigstim}$  component of the information and that the effects of the average level of correlation ( $I_{corind}$ ) and the stimulus dependence of correlation ( $I_{cordep}$ ) cancel each other.

The intuition provided by this example pair applies generally across the data set. For the whole population at a time window of 5 ms, there was an average 3% redundancy. This redundancy arose despite the fact that correlation was contrast-dependent; the average contribution of the stimulus-dependent correlation component  $I_{cordep}$  was 6.3% of the total information (Fig. 8*A,F*). Although the overall level of correlation (Fig. 8*C*,  $I_{cor}$ ) covaried significantly across the data set with synergy fraction,  $I_{cordep}$  did not (Fig. 8*D*). Much of the correspondence between  $I_{cor}$  and

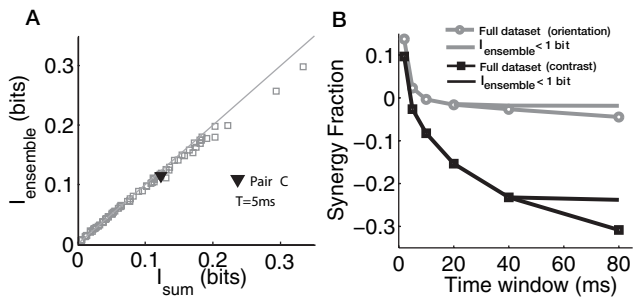


**Figure 8.** Contrast coding across the population of 71 pairs of cells for which contrast data were available. *A*, Histogram of the relative contribution of  $I_{cordep}$  to the total information for a time window of 5 ms. *B*, The relationship between  $I_{sigstim}$ , the component reflecting the contribution caused by the overlap in tuning curves, and the extent of redundancy as indicated by the negative synergy fraction. *C–E*, The relationship between the magnitude of each of the correlation components of the information and the overall extent of synergy on a pair-by-pair basis across the entire data set. *F*, A histogram showing the average fraction of information captured by each component across the data set. Error bars indicate SEM.

synergy/redundancy is accounted for by  $I_{corind}$ , which is strongly related to the synergy fraction ( $r = +0.58$ ;  $p = 1 \times 10^{-5}$ ) (Fig. 8*E*). The effect of the correlation components for contrast is relatively modest. Figure 8, *B* and *F*, shows that it is  $I_{sigstim}$  that most strongly predicts the level of redundancy ( $r = +0.62$ ;  $p = 1 \times 10^{-5}$ ), indicating that the redundancy substantially arises from the high degree of similarity of all contrast-tuning curves (compare Figs. 5*B*, 8*B*, the differing nature of the role of  $I_{sigstim}$  in contrast and direction coding).  $I_{cordep}$  nevertheless does have an effect on the coding regimen, in that its lower value for the contrast data set than for the direction data set means that, for contrast, the redundant effects are not compensated for (compare Figs. 5*F*, 8*F*). These results in general are affected by the time window used to count spikes. The information is maximal (and redundancy lowest) at  $T = 2$  ms, a time scale that corresponds to a strong effect of fine time scale synchronization. As the time window is increased and the effect of spike locking is averaged out, the redundancy increases, reaching a maximal value at 80 ms.

The relative uniformity of the redundancy for the contrast data set is apparent from Figure 9*A*. Figure 9*B* summarizes the results over the population for the coding of contrast compared with direction coding. The pairs are weakly redundant at 5 ms; however, by 40 ms time windows this has turned into strong





**Figure 9.** Redundancy is observed across the entire data set for contrast coding. *A*, Even at  $T = 5$  ms, redundancy was observed (points below the diagonal). The pair marked C is that referred to in Figure 7. *B*, The degree of redundancy in the coding of contrast quickly becomes substantial as the time window is increased. The results for contrast coding are indicated by squares and a black line; the results for direction coding (from Fig. 3*B*) are indicated by the gray line and circles.

redundancy (synergy fraction equal to  $-0.235 \pm 0.005$ ). Our findings show that the coding for contrast is more redundant than for direction, regardless of the window of analysis and despite the presence of significant correlation. This is consistent with the general picture that pairs of cells with similar tuning (typically the case for monotonic contrast-tuning functions and true for a subset of the direction tuning data) have high correlation but that the coding is mainly redundant.

## Discussion

We found that, for direction coding, the information available by being precise about spike timing for pairs of neurons in V1 adds superlinearly, but as the time scale becomes more coarse, information about direction adds slightly sublinearly across pairs of cells, in agreement with a previous information theoretic study of V1 cells (Reich et al., 2001). We have revealed the origin of these effects by our information component analysis, which shows a redundant (negative) contribution to the information attributable to the overlap in tuning and to the average level of correlation (i.e., stimulus-independent correlation). However, these components are balanced on average by a contribution from stimulus-dependent synchrony.

Synergy is stronger (and redundancy weaker) for pairs of neurons with dissimilar tuning curves. Pairs of neurons with overlapping spatial receptive fields but dissimilar orientation tuning will provide a strong joint response to features such as corners and T-junctions in a visual scene (Das and Gilbert, 1999); by comparison, those pairs with similar orientation tuning will be jointly driven best by lines and edges. Our results suggest that the pairwise coding of lines and edges is essentially independent, attributable to the effects of tuning-related redundancy, whereas the coding of features such as corners should be more synergistic. Analogously, the coding for changes in contrast is strongly redundant, attributable to the greater inherent similarity of the contrast-tuning curves of the neurons in a pair.

The fact that the synchronization depends on the stimulus direction is critically important for coding. It serves to reduce the redundancy caused by cells having similar tuning. In our experiments we used direction stimuli that were relatively coarsely spaced (typically  $22.5^\circ$  increments). Another study used finer increments of orientation (Samonds et al., 2003) and found the degree of fine time scale spike synchronization to be exquisitely dependent on stimulus orientation. Our study therefore might underestimate the contribution of stimulus-dependent synchrony to the neural population code for direction/orientation.

Contrast response functions of pairs of neurons tend to be fairly similar, and synchrony at lower contrasts is less temporally precise. For both contrast and direction coding the redundant contributions are dominated by the similarity in tuning, although the redundancy is stronger for contrast. The average level of correlation plays a small redundant effect in both cases. The synergistic contribution of the stimulus dependence of correlations is also lower on average for contrast than for direction coding. Overall, this results in a coding regimen more similar to the traditional intuition, in which correlations tend to result in redundancy and thus limit the number of neurons for which the outputs could be combined usefully to represent the stimulus variable. Importantly, the results of our analysis warn strongly against making general statements about the role of correlations in neural coding; as demonstrated here, the effect of correlations can be quite sensitive to both the time scale and the nature of the stimulus parameter that is being studied.

These principles underlying the neural coding of stimulus orientation and contrast were revealed by performing, for the first time, an information component breakdown of V1 neuronal responses. In doing so, it was crucial to use an effective sampling procedure (Nemenman et al., 2004) to avoid the results being contaminated by residual bias. The sampling approach we took is described in detail in supplemental material A (available at [www.jneurosci.org](http://www.jneurosci.org)). The information component breakdown rendered transparent the interplay of synergistic and redundant interactions, giving rise to the overall result of weakly synergistic coding at short time scales and independent orientation (and strongly redundant contrast) coding at longer time scales. The information components, first derived for short time windows in Panzeri et al. (1999) and extended to arbitrary time scales by Pola et al. (2003), relate to terms that have been measured by others (for review, see Schneidman et al., 2003). In particular, the stimulus-dependent correlation component,  $I_{\text{cordep}}$ , is exactly equal to the quantity  $\Delta I$  computed by Nirenberg et al. (2001) and Nirenberg and Latham (2003).

Schneidman et al. (2003) raised the importance of the distinction among different kinds of independence in neural coding. A first type of independence (which we will call type I) is activity or response independence, which might apply if spike trains truly were uncorrelated, i.e.,  $p(r_1, r_2) = p(r_1)p(r_2)$ . A second kind is conditional independence, which might apply if correlated activity could be explained purely by overlap in receptive field properties. This type II independence notion accounts for the oftmade distinction between “signal” and “noise” correlations: only the former involve type II dependence. Type III independence is informational independence and applies to the situation when information adds linearly across cells, possibly despite the presence of activity and/or conditional dependencies. This latter case describes the situation for our direction data set at long time scales and is a fair description of the situation at short time scales given the weak synergy we observe. However, it is important to note that even a small amount of pairwise informational dependence can have a relatively large effect on the population code for direction (Averbeck et al., 2006; Schneidman et al., 2006; Shlens et al., 2006), even if only pairwise correlations are present. Such an effect can be expected to be even larger if the pairwise correlations are themselves only the second order signature of higher order correlations. Thus deviations from informational independence of the order of 10% may well be of substantial importance.

The primary goal of the current study was to address how synchronization of spike trains affects the neural coding of drift-gratings of different directions and contrasts. This leads to an

obvious question: what is synchronization for? A number of the theories have been proposed, including temporal binding (Singer and Gray, 1995), fine orientation discrimination (Samonds et al., 2003, 2004), and contour integration (Samonds et al., 2006). Our results suggest a new role for stimulus-dependent synchronization: to create a coding regimen that allows sensory information to be pooled across a neuronal population in a linear (i.e., informationally independent) or even superlinear way despite the presence of redundancy in the signals conveyed by individual cells (i.e., their similarity in tuning). This may be an important role, particularly if the redundancy is a necessary feature of cortical circuitry rather than an imperfection. Because redundancy can lead to improved robustness through fault tolerance, it may well be desirable. By balancing this redundancy with stimulus-dependent synchronization, the brain could take advantage of both the robustness provided by redundancy and the accuracy attributable to pooling allowed by informational independence.

We find that destroying the identity of which neuron fired each action potential results in a substantial loss in information about stimulus direction. This provokes the following question: how could downstream neurons decode and make use of the positive information contribution provided by the stimulus-dependent synchronization? Simple linear readout schemes such as the population vector (Georgopoulos et al., 1986) cannot; a nonlinear readout algorithm is required, such as the nonlinear population vector (Shamir and Sompolinsky, 2004). This does not address how such a readout should be implemented, however, and the implementation constraints are particularly severe if one considers that the information should be usable within a single neuronal layer, as opposed to a multi-layer network. An integrate-and-fire operation effectively would pool input spikes regardless of origin, thus being subject to the Zohary et al. (1994) limitations on combining information from correlated input neurons. Nonlinear dendritic summation (Häusser and Mel, 2003) thus would appear to be a necessary feature for the biophysical implementation of a decoder capable of making use of the additional information contribution.

## References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.
- Averbeck BB, Lee D (2003) Neural noise and movement-related codes in the macaque supplement area. *J Neurosci* 23:7630–7641.
- Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358–366.
- 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.
- Cavanaugh JR, Bair W, Movshon JA (2002) Nature and interaction of signals from the receptive field center and surround in macaque V1. *J Neurophysiol* 88:2530–2546.
- Cover TM, Thomas JA Elements of information theory. New York: Wiley.
- Dan Y, Alonso J-M, Usrey WM, Reid RC (1998) Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci* 1:501–507.
- Das A, Gilbert CD (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature* 399:655–661.
- De Angelis GC, Ghose GM, Ohzawa I, Freeman RD (1999) Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J Neurosci* 19:4046–4064.
- De Charms RC, Merzenich MM (1996) Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381:610–613.
- De Oliveira SC, Thiele A, Hoffman KP (1997) Synchronization of neural activity during stimulus expectation in a direction discrimination task. *J Neurosci* 17:9248–9260.
- Engel AK, König P, Gray CM, Singer W (1990) Stimulus-dependent neuronal oscillations in cat visual cortex: inter-columnar interactions as determined by cross-correlation analysis. *Biol Cybern* 60:121–130.
- Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci* 13:2758–2771.
- Gawne TJ, Kjaer TW, Herz JA, Richmond BJ (1996) Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cereb Cortex* 6:482–489.
- Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 243:1416–1419.
- Golledge HD, Panzeri S, Zheng F, Pola G, Scannell JW, Giannikopoulos DV, Mason RJ, Tovee MJ, Young MP (2003) Correlations, feature-binding and population coding in primary visual cortex. *NeuroReport* 14:1045–1050.
- Gray CM (1999) The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron* 24:31–47.
- Häusser M, Mel B (2003) Dendrites: bug or feature? *Curr Opin Neurobiol* 13:372–383.
- Imas OA, Ropella KM, Wood JD, Hudetz AG (2004) Halothane augments event-related gamma oscillations in rat visual cortex. *Neuroscience* 123:269–278.
- Kaysner C, König P (2004) Population coding of orientation in the visual cortex of alert cats—an information theoretic analysis. *NeuroReport* 15:2761–2764.
- Kohn A, Smith MA (2005) Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J Neurosci* 25:3661–3673.
- Kreiter AK, Singer W (1992) Oscillatory neuronal responses in the visual cortex of awake macaque monkey. *Eur J Neurosci* 4:369–375.
- Kreiter AK, Singer W (1996) Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. *J Neurosci* 16:2381–2396.
- Latham PE, Nirenberg S (2005) Synergy, redundancy, and independence in population codes, revisited. *J Neurosci* 25:5195–5206.
- Lebedev MA, Mirabella G, Erchova I, Diamond ME (2000) Experience dependence plasticity of rat barrel cortex: redistribution of activity across barrel-columns. *Cereb Cortex* 10:23–31.
- Lee D, Port NL, Kruse W, Georgopoulos AP (1998) Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *J Neurosci* 18:1161–1170.
- Maldonado PE, Friedman-Hill S, Gray CM (2000) Dynamics of striate cortical activity in the alert macaque. II. Fast time scale synchronization. *Cereb Cortex* 10:1117–1131.
- Mastrorarde DN (1983) Correlated firing of cat retinal ganglion cells. I. Spontaneously active input to X- and Y-cells. *J Neurophysiol* 49:303–324.
- Mazurek ME, Shadlen MN (2002) Limits to the temporal fidelity of cortical spike rate signals. *Nat Neurosci* 5:463–471.
- Meister M (1996) Multineuronal codes in retinal signaling. *Proc Natl Acad Sci USA* 93:609–614.
- Movshon JA, Albright TD, Stoner GR, Majaj NJ, Smith MA (2003) Cortical responses to visual motion in alert and anesthetized monkeys. *Nat Neurosci* 6:3.
- Nelson JJ, Salin PA, Munk MH-J, Bullier J (1992) Spatial and temporal coherence in cortico-cortical connections: a cross-correlation study in areas 17–18 in the cat. *Vis Neurosci* 9:21–37.
- Nemenman I, Shafee F, Bialek W (2002) Entropy and inference, revisited. In: *Advances in neural information processing 14* (Dietterich TG, Becker S, Ghahramani Z, eds). Cambridge, MA: MIT.
- Nemenman I, Bialek W, de Ruyter van Steveninck R (2004) Entropy and information in neuronal spike trains: progress on the sampling problem (E-pub). *Phys Rev E Stat Nonlin Soft Matter Phys* 69:056111.
- Nirenberg S, Latham PE (2003) Decoding neuronal spike trains: how important are correlations? *Proc Natl Acad Sci USA* 100:7348–7353.
- Nirenberg S, Carciari SM, Jacobs AL, Latham PE (2001) Retinal ganglion cells act largely as independent encoders. *Nature* 411:698–701.
- Nowak LG, Munk MHJ, James AC, Girard P, Bullier J (1999) Cross correlation study of the temporal interactions between areas V1 and V2 of the macaque monkey. *J Neurophysiol* 81:1057–1074.
- Oram MW, Hatsopoulos NG, Richmond BJ, Donoghue JP (2001) Excess synchrony in motor cortical neurons provides redundant direction information with that from coarse temporal measures. *J Neurophysiol* 86:1700–1716.

- Panzeri S, Schultz SR (2001) A unified approach to the study of temporal, correlational, and rate coding. *Neural Comput* 13:1311–1349.
- Panzeri S, Treves A (1996) Analytical estimates of limited sampling biases in different information measures. *Network* 7:87–107.
- 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.
- Panzeri S, Petersen SR, Schultz SR, Lebedev M, Diamond M (2001) The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* 29:769–777.
- Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys J* 7:419–440.
- Petersen RS, Panzeri S, Diamond ME (2001) Population coding of stimulus location in rat somatosensory cortex. *Neuron* 32:503–514.
- Pola G, Thiele A, Hoffmann K-P, Panzeri S (2003) An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network* 14:35–60.
- Puchalla JL, Schneidman E, Harris RA, Berry MJ (2005) Redundancy in the population code of the retina. *Neuron* 46:493–504.
- Reich DS, Mechler F, Victor JD (2001) Independent and redundant information in nearby cortical neurons. *Science* 294:2566–2568.
- Richmond BJ, Gawne TJ (1998) The relationship between neuronal codes and cortical organization. In: *Neuronal ensembles, strategies for recording and decoding* (Eichenbaum HB, Davis JL, eds), pp 57–79. New York: Wiley.
- Roelfsema PR, Engel AK, Koenig P, Singer W (1997) Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* 385:157–161.
- Samonds JM, Bonds AB (2005) Gamma oscillation maintains stimulus structure-dependent synchronization in cat visual cortex. *J Neurophysiol* 93:223–236.
- Samonds JM, Allison JD, Brown HA, Bonds AB (2003) Cooperation between area 17 neuron pairs enhances fine discrimination of orientation. *J Neurosci* 23:2416–2425.
- Samonds JM, Allison JD, Brown HA, Bonds AB (2004) Cooperative synchronized assemblies enhance orientation discrimination. *Proc Natl Acad Sci USA* 101:6722–6727.
- Samonds JM, Zhou Z, Bernard MR, Bonds AB (2006) Synchronous activity in cat visual cortex encodes collinear and cocircular contours. *J Neurophysiol* 95:2602–2616.
- Schneidman E, Bialek W, Berry MJ (2003) Synergy, redundancy, and independence in population codes. *J Neurosci* 23:11539–11553.
- Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
- Schultz SR, Panzeri S (2001) Temporal correlations and neural spike train entropy. *Phys Rev Lett* 86:5823–5826.
- Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* 18:3870–3896.
- Shamir M, Sompolinsky H (2004) Nonlinear population codes. *Neural Comput* 16:1105–1136.
- Shannon CE (1948) A mathematical theory of communication. *Bell System Tech J* 27:379–423.
- Shlens J, Field G, Gauthier J, Grivich M, Petrusca D, Sher A, Litke A, Chichilnisky EJ (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26:8254–8266.
- Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18:555–586.
- Tolhurst DJ (1989) The amount of information transmitted about contrast by neurons in cat's visual cortex. *Vis Neurosci* 2:409–413.
- Ts'o DY, Gilbert C (1988) The organization of chromatic and spatial interactions in the primate striate cortex. *J Neurosci* 8:1712–1727.
- Werner G, Mountcastle VB (1965) Neural activity in mechanoreceptive cutaneous afferents: stimulus–response relations, Weber functions, and information transmission. *J Neurophysiol* 28:359–397.
- Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.