



Neural Population Structures and Consequences for Neural Coding

DON H. JOHNSON

*Department of Electrical & Computer Engineering, MS 366, Rice University, 6100 Main Street, Houston, Texas,
TX 77251-1892, USA*

dhj@rice.edu

Received September 14, 2002; Revised May 20, 2003; Accepted October 3, 2003

Action Editor: Terrence Sejnowski

Abstract. Researchers studying neural coding have speculated that populations of neurons would more effectively represent the stimulus if the neurons “cooperated:” by interacting through lateral connections, the neurons would process and represent information better than if they functioned independently. We apply our new theory of information processing to determine the fidelity limits of simple population structures to encode stimulus features. We focus on noncooperative populations, which have no lateral connections. We show that they always exhibit positively correlated responses and that as population size increases, they perfectly represent the information conveyed by their inputs regardless of the individual neuron’s coding scheme. Cooperative populations, which do have lateral connections, can, depending on the nature of the connections, perform better or worse than their noncooperative counterparts. We further show that common notions of synergy fail to capture the level of cooperation and to reflect the information processing properties of populations.

Keywords: neural coding, information theory, neural populations

1. Introduction

With advances in multi-electrode recording techniques and spike-sorting algorithms, researchers can discern the individual discharge patterns of several neurons located within a small region. Such recordings allow study of population coding: how does the population (neural assembly) as a whole represent information? Perhaps the simplest population occurs in sensory systems, where several neurons receive innervation from cells representing the same feature space. For example, in the peripheral auditory system, many auditory-nerve fibers innervate the same region of the cochlea, which means that their outputs represent the same frequency range. The populations receiving auditory-nerve fiber input would fall into the category of populations of concern here. The standard presumption is that neurons within a population “cooperate” to better encode information contained in their inputs.

Many investigators have developed statistical tests to determine whether such population coding is occurring or not. Most of these tests focus on *synergy* as measured by information theoretic quantities (entropy and mutual information): the population’s response considered as a whole represents stimuli better than the sum of individual neuron response contributions (Alonso et al., 1996; Bezzi et al., 2001; deCharms and Merzenich, 1996; Kang and Sompolinsky, 2001; Kilgard and Merzenich, 1999; Nirenberg et al., 2001; Reich et al., 2001; Warland et al., 1997).

Surveying the literature reveals remarkably inconsistent findings; some report synergy, some none, and some anti-synergy (the whole conveys less than the sum of individual contributions) in population recordings. In this paper, we endeavor to understand from a general theoretical perspective how well the simple neural population structures shown in Fig. 1 can represent the stimulus information expressed by their

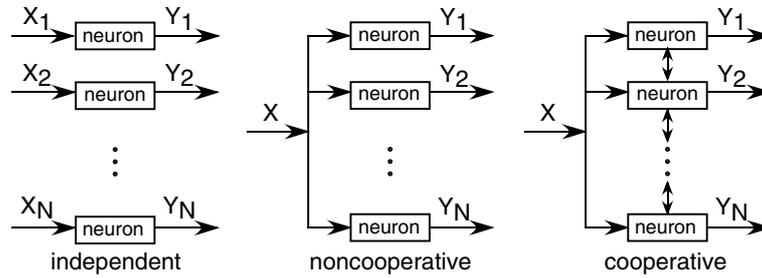


Figure 1. A neural population is defined here as a group of statistically identical neurons that receive a common set of inputs X . “Statistically identical” means that the probability distribution of the output for a given stimulus is the same for each neuron. In such homogeneous populations, the responses will differ in detail but each has the same average response. The simplest population, the independent one shown at the left, has statistically identical and independent inputs, which makes the outputs statistically independent. The noncooperative configuration has neurons with a common input but no other interactions, and thus reflects stimulus-induced dependencies. In the cooperative case shown on the right, the neurons interact and thus express stimulus- and connection-induced dependencies. For graphic simplicity, only nearest-neighbor interactions are shown; in general, any kind of interactions are allowed.

inputs. The framework for this analysis is our new theory of information processing (Johnson et al., 2001; Sinanović and Johnson, 2004). This theory quantifies how well single- and multi-neuron discharge patterns encode information according to a coding fidelity measure. Rather than using entropy or mutual information, we define synergy with our measure in a way similar to previous definitions. We use our coding fidelity measure to quantify how well systems can process information by operating on their inputs and encoding the result in a population code. Other than requiring that neural discharge patterns have a probabilistic description, our theory makes no assumptions about the neurons, what stimulus features are encoded, or what the neural code is. Because we do not rely on neural models to make theoretical predictions, the source of discharge pattern randomness is not specified: it could arise from the stimulus, the neural input, from the spike generator, or any combination of these. Furthermore, we are not restricted to analytic models chosen for tractability rather than appropriateness (e.g., Gaussian models). Rather than use detailed statistical models for the population’s joint statistics, we focus on the information processing properties inherent to various population structures. Using our theory, we show that simple cooperative populations, which contain both stimulus-induced and connection-induced statistical dependencies among the individual neural outputs, can express both synergy (the population expresses the stimulus better than the sum of individual neuron contributions) and anti-synergy (the population expresses the stimulus less well than the sum). Which occurs depends on the statistical relationships among the neural responses that result from the cooperation. Regardless of which

measure is used, ours or others, we show that synergy measures do not assess cooperation and that structures that do not exhibit synergy can encode stimuli as well as theoretically possible. In particular, we show that noncooperative populations, which only have stimulus-induced dependencies, can *always* represent stimulus features with arbitrary fidelity regardless of the neural code if the population size is sufficiently large.

2. Methods

We use the symbols X and Y to denote a neuron’s input and output, respectively (Fig. 1). What aspects of the input and output code information need not be stated in our theory. For example, X could represent the spike count over some time interval, spike occurrence times, or some other quantity extracted from the spike train. The input X could also express a population’s response by representing the spike times of several neurons, and even incorporate the state of some neuromodulator. The population output is expressed by the vector $\mathbf{Y} = [Y_1, \dots, Y_N]$, with Y_n representing the n th neuron’s output and N denoting the population size. As with the input, Y_n can represent any aspect of the discharge pattern. Because of this generality and because we don’t rely on any neural models, the results contained here apply to a wide variety of signaling mechanisms. The conditional probability function $p_{Y_n|X}(y_n | x)$ describes each neuron’s input-output relationship in statistical terms. When this conditional probability function is the same for each neuron, we have a *homogeneous* population. In this case, each neuron’s output obeys the same probability law as the others, but the output signals will differ.

Our work focuses on the three population structures shown in Fig. 1. When each neuron's input is statistically independent of the other neurons' inputs and the neurons do not cooperate, we have the *independent* population. In an independent structure, the individual neuron responses are statistically independent of each other and, as we show later, this population exhibits no information processing gain beyond that of a single neuron. In a *noncooperative* population, the common innervation of population neurons means that each neuron's response is statistically dependent on the others to some degree; we refer to this dependence as *stimulus-induced dependence*.¹ In a *cooperative* population, the neurons interact with each other via collaterals and/or interneurons—cooperate—to create *connection-induced dependence*. In general, cooperative populations also have stimulus-induced dependence.

The three population structures are mathematically specified by the different forms they yield for $p_{\mathbf{Y}}(\mathbf{y}; \alpha)$ that describes the joint probability function of the aggregate population output. We assume that this probability function varies with the stimulus, represented here symbolically by α . For the independent structure, the stimulus drives a set of statistically independent inputs that serve a population of noncooperative neurons. The outputs are thus statistically independent, and the population's probability function factors: $p_{\mathbf{Y}}(\mathbf{y}; \alpha) = p_{Y_1}(y_1; \alpha) \cdot p_{Y_2}(y_2; \alpha) \dots p_{Y_N}(y_N; \alpha)$, which can be more succinctly expressed using the product notation $p_{\mathbf{Y}}(\mathbf{y}; \alpha) = \prod_{n=1}^N p_{Y_n}(y_n; \alpha)$. The noncooperative structure has the property that for each specific input, the population output has statistically independent components. "Specific input" here means a particular value for the neural input X , a specific sequence of spike times for example, but not some particular stimulus. Mathematically, this property means that the outputs are statistically independent when conditioned on the input: $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | x) = \prod_{n=1}^N p_{Y_n|\mathbf{X}}(y_n | x)$. The omission of α from terms in the product is intentional; this notation expresses the assumption that each neuron has no knowledge of the stimulus attribute other through its input X . The n th neuron's output probability function and the population's joint probability function are

$$p_{Y_n}(y_n; \alpha) = \int p_{Y_n|\mathbf{X}}(y_n | x) p_X(x; \alpha) dx \quad (1)$$

$$p_{\mathbf{Y}}(\mathbf{y}; \alpha) = \int \prod_{n=1}^N p_{Y_n|\mathbf{X}}(y_n | x) p_X(x; \alpha) dx \quad (2)$$

Because of the integral, the joint probability function does not factor and we have mathematically expressed stimulus-induced dependence. In the *cooperative* case, the joint probability distribution does not have any particular form, which means that both connection- and stimulus-induced dependence can occur.

Typically, theoretical analyses of neural coding employ Fisher information (Abbott and Dayan, 1999; Wilke and Eurich, 2002; Shamir and Sompolinsky, 2001) or measures widely used in information theory such as entropy and mutual information (Tishby et al., 1999; Strong et al., 1998; Warland et al., 1997). We use an information-theoretic distance here to tackle neural coding in a broader fashion. In our theory of information processing, which is described more fully in Sinanović and Johnson (2004), a neural code's efficacy can be judged by the ability to discriminate a stimulus change from the response and by the ability to estimate the stimulus from the response. We wanted a quantity related to the classification and estimation performance levels a hypothetical optimal processor could achieve by observing single-neuron or population responses. Fisher information is directly related to estimator performance through the Cramér-Rao bound (Abbott and Dayan, 1999; Johnson et al., 2001), but not to classification. Neither entropy nor mutual information can be directly related to either processing criterion.² Rate distortion theory offers a way to relate mutual information to classification and estimation issues in a very general way (Shannon, 1948; Kolmogorov, 1956; Berger, 1971). This theory forms the heart of the information bottleneck method (Tishby et al., 1999). However, this approach requires the stimulus to have a stochastic model and an explicitly specified measure of the error between the encoded and decoded stimulus. Choices for either the stimulus model and the error measure greatly affect the results (Johnson, 2002).

In addition to desiring a quantity that could assess both classification and estimation abilities, we wanted it to act like an "information probe." We wanted a measure that could be used to analyze any response measured along a pathway so that we could determine what stimulus features were well encoded at each stage and how well each neural system processed its input. We did not want to limit the kind of response studied; we wanted to encompass single-neuron discharge patterns, population discharges, intracellular signals, and behaviors. With this information probe, we could empirically or theoretically study how a pathway processed and integrated information. This work describes a theoretical

study of simple population structures that reaches general conclusions on how well a population's collective output represents information.

Our information probe is an information-theoretic distance between the probability distributions describing the responses to two stimulus conditions (Johnson et al., 2001). The distance measure we use is the Kullback-Leibler (KL) distance because it quantifies *coding effectiveness*: how well can the stimulus change be discerned from the responses and how well can the stimulus be determined. Details on the theoretical reasons for using the KL distance are found elsewhere (Sinanović and Johnson, 2004); empirical methods are detailed in Johnson et al. (2001). The KL distances between the population's input and output that result when the stimulus is α_0 and when it is α_1 are given by the following integrals.

$$D_X(\alpha_1 \parallel \alpha_0) = \int p_X(x; \alpha_1) \log \frac{p_X(x; \alpha_1)}{p_X(x; \alpha_0)} dx \quad (3)$$

$$D_Y(\alpha_1 \parallel \alpha_0) = \int p_Y(y; \alpha_1) \log \frac{p_Y(y; \alpha_1)}{p_Y(y; \alpha_0)} dy \quad (4)$$

The logarithm base is arbitrary; we use base-two logarithms, which makes our distance have units of bits. The *only* assumption made in information processing theory is that the signals are stochastic, and thus are described entirely by their probability functions. We make no assumption about the stimulus; it can be deterministic or stochastic. To calculate the KL distance, we need the probability model (either actual or measured) describing the response such as expressed by the probability functions in Eqs. (1) and (2). The relation of the KL distance to optimal classification and estimation is detailed in Johnson et al. (2001). Briefly, the error probability of an optimal classifier that tries to distinguish between the responses to the stimuli parameterized by α_0 and α_1 decreases exponentially in the KL distance: $P_e \sim 2^{-D(\alpha_1 \parallel \alpha_0)}$. Furthermore, for small differences between the stimuli, the KL distance is proportional to the Fisher information: $D(\alpha_0 + \delta\alpha \parallel \alpha_0) \approx F(\alpha_0) \cdot (\delta\alpha)^2/2$. Thus, with one quantity we can assess coding effectiveness from both classification and estimation perspectives.

The KL distance between random vectors having statistically independent components equals the sum of the distances between each component (Johnson et al., 2001): $D_Y(\alpha_1 \parallel \alpha_0) = \sum_{n=1}^N D_{Y_n}(\alpha_1 \parallel \alpha_0)$. This so-called additivity property is shared by the entropy and mutual information measures. For independent neural

populations, population coding effectiveness equals the sum of what individual neuron contribute. We define *synergy* (anti-synergy) to occur when the population distance is greater (less) than the sum of individual distances, a definition similar to that when entropy or mutual information measures are used. Thus, *the independent population defines the baseline against which we determine synergy*. We show later that populations can show synergy, anti-synergy, or both.

How well a population processes the stimulus information contained in its input is quantified by the *information transfer ratio* γ of KL distances between the population's responses to two stimulus conditions and between the population's inputs corresponding to the same stimuli (Johnson et al., 2001).

$$\gamma_{X,Y}(\alpha_0, \alpha_1) = \frac{D_Y(\alpha_1 \parallel \alpha_0)}{D_X(\alpha_1 \parallel \alpha_0)}$$

The information transfer ratio quantifies how a system—here, a neural population—affects the ability to discern the stimulus from its output relative to its input. Note that we can find the information transfer ratio even when the neural codes of the input and output differ; in fact, the nature of signals could differ. For example, the input could be a post-synaptic potential and the output a spike train (Johnson et al., 2000; Rozell et al., 2002). Because KL distances and the information transfer ratio can be applied to any signal, they provide a general means of analyzing the information processing of neurons and populations. A fundamental result from information theory, the data processing theorem or inequality (Cover and Thomas, 1991; Johnson et al., 2001), states that this ratio must be less than or equal to one. This result means that *no* system can enhance the effectiveness of stimulus coding beyond that of its input. Thus, the information transfer ratio quantifies the *fidelity* of the population's information processing. Consequently, the fidelity cannot be greater than one: a population cannot create a less noisy representation of the stimulus than found in its input. This important concept from our theory of information processing applies regardless of the population structure or the kinds of neurons in the population; they can even be inhomogeneous and encode the stimulus differently. A small value for γ means that little of the information about the two stimulus conditions expressed by the input appears in the neural output. A value approaching one means that nearly all of the information in the input about the stimulus change appears in the output. We could have the situation wherein the population suppresses

some stimulus features so as to enhance the representation of some other feature. Such *information filtering* is certainly allowed. We usually take α_0 to represent a reference stimulus condition and α_1 to represent stimulus changes with respect to this reference. The stimulus change $\alpha_0 \rightarrow \alpha_1$ could correspond to changing a suppressed or a desired feature, which presumably results in small or large values for γ , respectively. Our previous work has shown that information filtering can be dynamically expressed in neural responses (Johnson et al., 2001).

Our goal here is to study the three population structures defined here—the independent, the noncooperative and the cooperative—to determine whether they can represent information conveyed by their inputs. To study the fidelity problem, we seek the asymptotic behavior of the information transfer ratio. As population size increases, is the population collectively able to achieve perfect fidelity? Mathematically, does $\gamma(N) \rightarrow 1$ as $N \rightarrow \infty$ for some stimulus attribute change? Furthermore, we want to determine how large the population needs to be to achieve a given fidelity and how this size depends on population structure. This approach allows a population's *redundancy* to be defined. In communications engineering, redundancy means more is available than needed to achieve some given level of fidelity. In our framework, wherein fidelity is quantified by the information transfer ratio, any population larger than that needed to achieve a specified value γ_0 is a redundant population. Whether a population can exhibit redundancy depends on its structure in addition to its size. For example, increasing independent and noncooperative populations results in the same structure. Studying redundancy in cooperative structures can be tricky. If we assume that cooperative populations have a regular structure, increasing the number of neurons elaborates the same underlying structure, making it easy to characterize its properties. However, the loss of any neuron in a regular structure may not result in a smaller population having the same structure. For independent and noncooperative structures, loss of any neuron leaves the structure unchanged.

3. Results

3.1. Correlation Between Neuron Outputs

As a preliminary, we need to characterize the stimulus-induced dependence found in homogeneous nonco-

operative populations. We use here the pairwise correlation coefficient ρ because of its simplicity. In Appendix A we show that this correlation coefficient cannot be negative. Thus, *stimulus-induced dependence in a homogeneous population always results in positive correlation between pairs of neural outputs: $\rho \geq 0$* . Maximal correlation ($\rho = 1$) occurs only in the noncooperative structure when the each neuron's output equals its input: the neurons introduce no randomness and serve as ideal relay cells. The correlation decreases with increasing noninput-related randomness in each neuron's output. Minimal correlation ($\rho = 0$) occurs when we have the independent structure or when each neuron's output does not depend on the input.

A simple example shows that if a population were to encode binary stimulus attributes (is the stimulus on or off, for example), synergy can result from correlation (the KL distance obtained at the population's output can be greater than that provided by the independent population). We considered a two-neuron population, each of which is described by a Bernoulli model for the number of spikes occurring in a time bin. This model merely says that no more than one spike can occur in a bin and that the probability of a spike in a bin is p , which equals the product of discharge rate and binwidth. The binwidth is usually chosen so that the resulting probability is never large (does not exceed 0.5, for example).

$$\Pr[m \text{ spikes in a bin}] = \begin{cases} p & m = 1 \\ 1 - p & m = 0 \\ 0 & m > 1 \end{cases} \quad (5)$$

Bernoulli models for the two-neuron population are specified by the correlation coefficient ρ and the spike probabilities for each neuron.³

$$\begin{aligned} & \Pr[Y_1 = m_1; Y_2 = m_2] \\ &= \begin{cases} \rho\sqrt{p_1(1-p_1)p_2(1-p_2)} + p_1p_2 & m_1 = 1, m_2 = 1 \\ p_1 - \Pr[Y_1 = 1; Y_2 = 1] & m_1 = 1, m_2 = 0 \\ p_2 - \Pr[Y_1 = 1; Y_2 = 1] & m_1 = 0, m_2 = 1 \\ 1 - p_1 - p_2 + \Pr[Y_1 = 1; Y_2 = 1] & m_1 = 0, m_2 = 0 \end{cases} \end{aligned} \quad (6)$$

We assume that the spike probabilities and the correlation can vary with stimulus attribute. Note that because we model the joint probability function of the population's output, we do not directly specify what population structures can yield various combinations

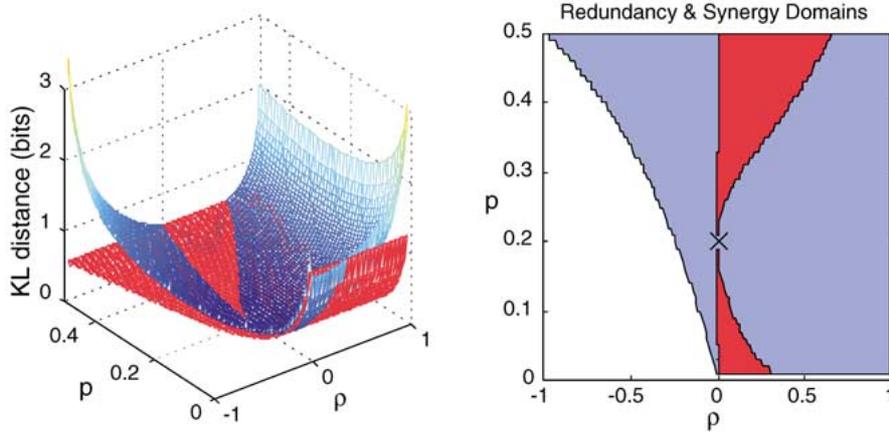


Figure 2. We used a Bernoulli model for a simulated two-neuron population. The first stimulus condition produced a spike probability of $p = 0.2$ and a correlation $\rho = 0$. The second stimulus condition yielded values of p and ρ located somewhere in their allowable domain (nonallowable combinations of p , ρ yield negative probabilities for the population's joint response). The left panel shows the Kullback-Leibler distances (red) achieved by the population (color-coded) and by the sum of the individual neuron's distances (red). The red-filled region in the right panel corresponds to the domain over which anti-synergy occurs; the blue-filled regions denote the domain in which synergy resulted. Those values of p and ρ not colored are nonallowable. The \times marks the probability-correlation value corresponding to the first stimulus. Large synergies result if the neurons are correlated in the correct way.

of p and ρ . We have shown that noncooperative populations can only exhibit positive correlations; cooperative populations can exhibit positive or negative correlations. For the reference stimulus condition in this example, we assume the population respond with identical spike probabilities, equal to p_0 , and are uncorrelated ($\rho = 0$). In this model, the $\rho = 0$ case corresponds to the independent model and serves as the synergy baseline. For the second stimulus condition, both neurons respond with the same spike probability (and thus remain a homogeneous population) and can become correlated. By computing the KL distance between this baseline and the correlated response for a variety of spike probability and correlation values, we see in Fig. 2 that both synergy and anti-synergy can result. The region of spike probability and correlation values that yield anti-synergy is quite limited and the difference of the population KL distance from the baseline value in this region is small. Be that as it may, noncooperative population structures, which exhibit positive correlation, can lie in this region. Outside of this region, which constitutes most of the feasible range of p and ρ , synergy occurs and the distances achieved in this region can be large. The region corresponding to negative correlation values has generally larger distances than in the positive correlation region, indicating that cooperative populations can demonstrate greater KL distances than noncooperative ones. Consequently, co-

operation can potentially enhance somewhat the ability of this simple two-neuron population to represent stimulus change. The synergy region can be expressed by both cooperative and noncooperative structures, and certainly this region corresponds to a better expression of the binary nature of the stimulus.

3.2. Effect of Population Size on Fidelity

We study population coding fidelity by analyzing how the information transfer ratio depends on population size. Considering the baseline independent structure first, wherein each neuron's input is statistically independent of other neuron inputs. Thus, the population outputs Y_n , $n = 1, \dots, N$, are also statistically independent of each other. The additivity property applies to both the input and the output.

$$\gamma_{\mathbf{X}, \mathbf{Y}}(\alpha_0, \alpha_1) = \frac{\sum_{n=1}^N D_{Y_n}(\alpha_0 \| \alpha_1)}{\sum_{n=1}^N D_{X_n}(\alpha_0 \| \alpha_1)}$$

Consequently, when the population is homogeneous, *the information transfer ratio is a constant*, equaling that of an individual neuron $\gamma_{\mathbf{X}, \mathbf{Y}}(\alpha_0, \alpha_1) = \gamma_{X_n, Y_n}(\alpha_0, \alpha_1)$ regardless of the population size. Simple bounding arguments reveal that in the general case

the overall information transfer ratio cannot exceed the largest individual information transfer ratio.

$$\gamma_{X,Y}(\alpha_0, \alpha_1) \leq \max_n \gamma_{X_n, Y_n}(\alpha_0, \alpha_1)$$

Thus, from an information processing viewpoint, the independent population is ineffective: one neuron could serve the purpose of many and only if each neuron has maximum fidelity ($\gamma_{X_n, Y_n}(\alpha_0, \alpha_1) = 1$ for all n) can perfect population fidelity be achieved. The independent population is highly *redundant*: more neurons are present than required to yield a given information transfer ratio. Thus, any of several neurons could be removed from this structure without affecting the population's ability to represent input information.

In contrast, we show in Appendix B that *any* noncooperative population, homogeneous or not, having dependent outputs can ultimately represent its input information *perfectly*. As population size N increases, the information transfer ratio monotonically increases ($\gamma(N+1) > \gamma(N)$) and that the information transfer ratio approaches one ($\lim_{N \rightarrow \infty} \gamma(N) = 1$). This result does not depend on each neuron's input code, how each neuron processes its input, or on the code each neuron uses to express the processed input in its output. This result contrasts with the independent population's behavior, which shows no variation in the information transfer ratio.

What is missing from this analysis of the noncooperative population is some idea how the information transfer ratio increases with population size. As shown in by Sinanović and Johnson (2004), the way γ increases with the population size for large populations depends on the way the *input* encodes the stimulus attribute. If the input uses a rate code (only the number of spikes occurring in a time interval encodes the stimulus), the information transfer ratio approaches one exponentially. If the input encodes the stimulus in continuous variables, such as spike timing or latency, γ approaches one hyperbolically.

$$\gamma(N) \stackrel{N \rightarrow \infty}{\approx} \begin{cases} 1 - k_1 \exp\{-k_2 N\} & X \text{ is discrete-valued} \\ 1 - k/N & X \text{ is continuous-valued} \end{cases} \quad (7)$$

The constants k , k_1 , and k_2 depend on the nature of both the input and each neuron's coding scheme, but these asymptotic results apply *regardless* of the coding each

neuron uses. In the continuous-valued case, examples indicate that the formula $\gamma(N) \approx (1 + k/N)^{-1}$ approximates the actual information transfer ratio quite well. Asymptotically, this approximation has the same form as the second part of (7).

Noncooperative structures cannot exhibit synergy. In our noncooperative structure, the input distance component of the information transfer ratio is a constant. The synergy baseline demands that the output distance be proportional to N , which in this case means that $\gamma(N)$ must increase linearly: $\gamma(N) = N \cdot \gamma(1)$. A linear increase in $\gamma(N)$ is impossible simply because $\gamma(N) \leq 1$. The asymptotic formulas (7) mean that any sufficiently large noncooperative population *cannot* demonstrate synergy; in fact they only demonstrate anti-synergy.⁴ Thus, *synergy is not a necessary property for populations either to exhibit an information processing gain or to achieve perfect fidelity*.

Determining the exact form of $\gamma(N)$ for any particular example and for small populations is difficult because the stimulus-induced statistical dependence among the population outputs makes evaluation of (2) difficult. Examples we can analyze are contrived from a neural coding viewpoint, but they illustrate how the constants in the asymptotic formulas depend on the coding parameters of individual neurons. In particular, let the input be a scalar random variable having an exponential distribution and the conditional probability function $p_{Y_n|X}(y_n | x)$ of the n th neuron's output describe a Poisson random variable.

$$\begin{aligned} p_X(x; a) &= a e^{-ax}, \quad x \geq 0 \\ p_{Y_n|X}(y_n | x) &= \frac{(G_n x)^{y_n} e^{-G_n x}}{y_n!}, \quad y_n = 0, 1, \dots \end{aligned} \quad (8)$$

In this model, the input determines the average number of Poisson-distributed spikes measured in the output over some time interval. G_n is the gain between the input and the output average spike count. The population would be homogeneous if all the gains were equal. Using Eq. (1), the unconditional output probability function for each neuron is found to be geometric.

$$p_Y(y_n; a) = \frac{a}{a + G_n} \left(\frac{G_n}{a + G_n} \right)^{y_n}, \quad y_n = 0, 1, 2, \dots$$

The input and output KL distances can be calculated exactly from Eqs. (2)–(4) when we change the parameter

a from a_0 to a_1 .

$$D_X(a_1 \| a_0) = \log \frac{a_1}{a_0} + \frac{1}{\ln 2} \frac{a_0 - a_1}{a_1}$$

$$D_Y(a_1 \| a_0) = \log \frac{a_1}{a_0} + \left(1 + \frac{1}{a_1} \sum_{n=1}^N G_n\right) \times \log \left(\frac{a_0 + \sum_{n=1}^N G_n}{a_1 + \sum_{n=1}^N G_n}\right)$$

As the population size N increases, the information transfer ratio, given by the ratio of these two distances, attains a value of unity only if the sum of gains diverges: for $\gamma(N) \rightarrow 1$ as $N \rightarrow \infty$, we must have $\sum_{n=1}^N G_n \rightarrow \infty$. This situation occurs when we have a homogeneous population ($G_n = G$) and the gain is not zero. The sum of gains will not be infinite when the gains decrease quickly, which corresponds to additional neurons diminishingly expressing the input. The output KL distance has the hyperbolic form in Eq. (7) with $k = (a_1 - a_0)^2 / (a_1 \bar{G} D_X(a_1 \| a_0))$ and \bar{G} equaling the average of the gain parameters across the population. Figure 3 shows the exact and asymptotic analysis results. The empirical equation mentioned earlier $\gamma(N) \approx (1 + k/N)^{-1}$ closely matches the exact result. Analytic calculations show the correlation coefficient between pairs of neurons, given by Eq. (9), to be $\rho_{n_1, n_2} = \left(\frac{G_{n_1} - G_{n_2}}{a + G_{n_1} + G_{n_2}}\right)^{1/2}$, confirming

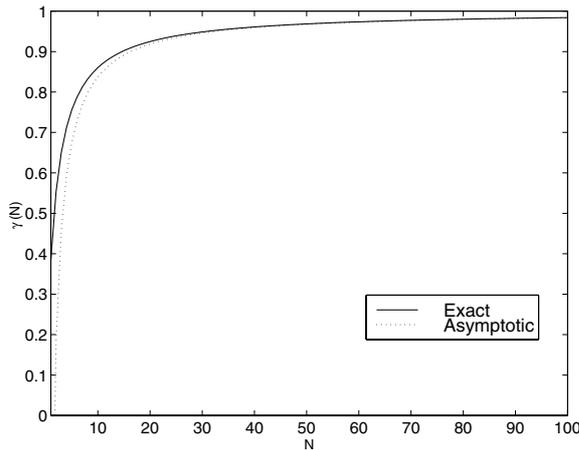


Figure 3. The information transfer ratio for the example population is plotted against the population size. The hyperbolic asymptotic formula is also shown as a dotted line. In this example, $a_0 = 2$, $a_1 = 1$, $G = 1$, which makes each neuron's information transfer ratio 0.38 and the correlation coefficient between neuron pairs equal to $1/3$ when $a = 1$.

that the correlation is indeed positive. Furthermore, this example *always* demonstrates anti-synergy because $\sum_{n=1}^N D_{Y_n}(a_1 \| a_0) > D_Y(a_1 \| a_0)$ for all $N > 1$. Yet, the information transfer ratio can come as close to one as desired when the population is large enough.

Changing the input in this Poisson example to a binary-valued random variable results in exponential rather than hyperbolic asymptotics, consistent with (7). To show this, let $X \in \{0, 1\}$ with $p_X(x = 0) = p$ and $p_X(x = 1) = 1 - p$. The input and output KL distances are

$$D_X(p_1 \| p_0) = p_1 \log \frac{p_1}{p_0} + (1 - p_1) \log \frac{1 - p_1}{1 - p_0}$$

$$D_Y(p_1 \| p_0) = \left(p_1 + (1 - p_1)e^{-\sum_n G_n}\right) \times \log \frac{p_1 + (1 - p_1)e^{-\sum_n G_n}}{p_0 + (1 - p_0)e^{-\sum_n G_n}} + (1 - p_1)(1 - e^{-\sum_n G_n}) \log \frac{1 - p_1}{1 - p_0}$$

Thus, if $\sum_{n=1}^N G_n \rightarrow \infty$, the information transfer ratio has an asymptotic value of one and approaches it exponentially. Consequently, a change in the nature of the *input* used by each neuron indeed changes the asymptotics of the population's information transfer ratio.

4. Conclusions

This paper addresses the fundamental information processing capabilities of admittedly simple population structures: the independent, the noncooperative, and the cooperative (Fig. 1). Our results are theories of structure, and do not depend on what neural codes the input or each constituent neuron employ: they apply to any collection of systems having these structures regardless of the signals they are processing and what signals they produce. Our most important result concerns the noncooperative structure: as the population grows, the information processing capability always increases to its theoretical limit. This result means that an ideal processing system could extract stimulus information from the population's collective output without incurring a fidelity loss. The purpose of the population can be more than serving as a relay; the population could be selecting some stimulus feature and suppressing others. Our results say that such information filtering for the selected feature can be accomplished with as perfect a fidelity as desired (relative to the input). This collective behavior occurs even when each neuron incurs information processing losses. This behavior represents the

kind of population coding researchers have envisioned; our results indicate that stimulus-induced dependence suffices to achieve perfect population coding fidelity.

The two different ways in which the information transfer ratio varies with population size depend entirely on the way the population's *inputs* encode the stimulus (Eq. (7)). It is tempting to say that discrete-valued inputs (i.e., rate codes) lead to "better" coding than do continuous-valued quantities because of the exponential increase of the information transfer ratio in the latter case. This temptation is mitigated by the fact this result is an asymptotic one, and only applies when the population is "large" (see Fig. 3, for example). Small populations may well exhibit different behaviors and ultimately the differences among input coding possibilities may not result in dramatically different information transfer ratios.

In addition to its information processing capability, the robustness of a structure to missing neurons needs to be considered. Rather than develop a separate measure of robustness, we considered how the information transfer ratio would change if neurons were removed from the structure. Coupled with this way of quantifying robustness is the notion of redundancy: how many neurons can be removed from the structure before its information processing performance falls below some critical level γ_0 . Noncooperative structures, including the independent structure, are robust to neuron loss, with the information transfer ratio changing in a predictable manner. The redundancy depends on the nature of the input code as well as the population size because of the differing asymptotic formulas for the information transfer ratio. The robustness and redundancy of cooperative structures depends on the nature of the neural interaction.

We have shown that synergy is not a sufficiently powerful notion to determine whether population coding employing cooperation is occurring. This conclusion does not hinge on our choice for the definition of synergy; rather, the problem with synergy is simply that it ignores the coding fidelity of the population input. Sufficiently large noncooperative populations can represent stimuli encoded in their inputs as accurately as desired without ever demonstrating synergy. In fact, the independent population that serves as the synergy baseline cannot represent information with perfect fidelity unless each neuron does. When we consider optimal rate coding of continuous-valued attributes over their entire range rather than binary-like stimulus changes (Johnson and Ray, 2004), we infer that populations hav-

ing statistically independent outputs, which have zero synergy, provide the best coding structure. We suggest there that cooperation in the form of a lateral connection network can produce independent outputs while achieving the same fidelity as a noncooperative one. In other words, connection-induced dependence can cancel stimulus-induced dependence. Those results might explain why so few experiments have found significant synergies in rate coding situations despite at least the suspicion that lateral connections exist.

We have not found general results for the improvements cooperative structures may have over noncooperative ones. Simple examples of cooperative, nonhomogeneous population structures illustrate how vital they can be for stimulus coding. To represent the direction a light bar is moving, interneuron timing differences might serve as an efficient code. Although we have not explored such models here, interesting cooperative structures would need to exhibit a greater information transfer ratio than their noncooperative counterparts: why have the added interconnections without providing better representation of information? Examining Fig. 3 indicates that cooperation cannot achieve large increases in the information transfer ratio. However, these increases could be important in that a given value of the information transfer ratio could be achieved with fewer neurons.

Whether the population has a cooperative structure or not, once a population is larger than some size, it reaches the point of diminishing returns: increases in the population do not result in significant information processing gains. The number of neurons needed to achieve a given value of $\gamma(N)$ depends on the information transfer ratios of individual neurons: the smaller they are, the more neurons are needed. For the homogeneous, noncooperative population processing a continuous-valued input, we can use the approximation $\gamma(N) \approx (1 + k/N)^{-1}$ to show that the constant $k = 1/\gamma(1) - 1$. Achieving a given value of $\gamma(N)$ requires the ratio k/N , which equals $1/(N\gamma(1)) - 1/N$, to be some value. Thus, a greater population size is needed to overcome neurons having small information transfer ratios. Once this critical size is achieved, further gains will be small and will require substantial increases in the population size. An information filtering perspective also suggests that populations be of limited size. Appendix B states that unless the individual information transfer ratios are zero, the information transfer ratio for a population approaches one monotonically as the population size increases. This result

means that any information filtering provided by each neuron will be washed out by the population unless the filtering is perfect. As our analysis suggests, the population size required to affect filtering properties depends on the information transfer ratio, with larger populations required the smaller the individual information transfer ratio. These considerations suggest that large populations are only needed when each neuron poorly processes its input, but if too large, the population response will diminish any differences in the information filtering of individual neurons.

Appendix A

The correlation coefficient between the n_1 th and n_2 th neurons in a homogeneous population equals

$$\rho_{Y_{n_1}, Y_{n_2}} = \frac{\mathcal{E}[Y_{n_1} Y_{n_2}] - (\mathcal{E}[Y])^2}{\mathcal{E}[Y^2] - (\mathcal{E}[Y])^2} \quad (9)$$

Because the population is homogeneous, Y denotes the output of any neuron and $\mathcal{E}[Y_{n_1}] = \mathcal{E}[Y_{n_2}] = \mathcal{E}[Y]$. To determine the range of values ρ can take on, we focus on the numerator and write its components using conditional expected values.

$$\begin{aligned} \mathcal{E}[Y_{n_1} Y_{n_2}] &= \mathcal{E}_X[\mathcal{E}[Y_{n_1} Y_{n_2} | X]] \\ &= \mathcal{E}_X[\mathcal{E}[Y_{n_1} | X] \mathcal{E}[Y_{n_2} | X]] \\ &= \mathcal{E}_X[(\mathcal{E}[Y | X])^2] \\ \mathcal{E}[Y] &= \mathcal{E}_X[\mathcal{E}[Y | X]] \end{aligned}$$

Here $\mathcal{E}_X[\cdot]$ means expected value with respect to the collective input's probability function. In going from the first to the second line, we used the fact that in non-cooperative structures, neural outputs are conditionally independent, which makes them conditionally uncorrelated. To prove our result, note that this expected value defines an inner product, $\mathcal{E}_X[f(X)g(X)] := \langle f(X), g(X) \rangle$. We apply the Cauchy-Schwarz inequality, which states for any inner product $\langle f(X), g(X) \rangle^2 \leq \langle f(X), f(X) \rangle \langle g(X), g(X) \rangle$. Equality occurs only when $f(X)$ is proportional to $g(X)$. We can show that the numerator of (9) is positive by letting $f(X) = \mathcal{E}[Y | X]$ and $g(X) = 1$. The Cauchy-Schwarz inequality becomes

$$(\mathcal{E}_X[\mathcal{E}[Y | X]])^2 \leq \mathcal{E}_X[(\mathcal{E}[Y | X])^2] \cdot \mathcal{E}_X[1]$$

The term $\mathcal{E}_X[1]$ equals one. Subtracting the left side from the right gives the numerator of (9) and we con-

clude that the numerator is always non-negative. The equality condition means that zero correlation only results when $\mathcal{E}[Y | X]$ equals a constant, which occurs only when the output's statistical characteristics do not vary with the input. Consequently, so long as each neuron's output reflects its input to some degree, the noncooperative population's outputs are positively correlated.

Because an independent population has each neuron receiving a separate input statistically independent of the other, the correlation coefficient is also zero. This result does not conflict with this derivation because X denotes the entire input and we assumed each neuron has access to it.

By comparing numerator and denominator of (9), we can understand the range of positive values the correlation coefficient can achieve. Because $\rho \leq 1$, simple rearrangement of (9) yields $\mathcal{E}_X[\mathcal{E}[Y^2 | X]] - \mathcal{E}_X[(\mathcal{E}[Y | X])^2] \geq 0$, which can also be written as $\mathcal{E}_X[\mathcal{E}[Y^2 | X] - (\mathcal{E}[Y | X])^2] \geq 0$. The quantity inside the outer brackets is the conditional variance of each neuron's output for a given input. For this conditional variance to equal zero, which would imply $\rho = 1$, the neuron output Y would need to equal a constant times the input X . This situation corresponds to the case where each neuron is a relay cell, which expresses its input in a noiseless way.

Appendix B

The limit achieved by the information transfer ratio for a homogeneous, noncooperative population is one, corresponding to perfect information transfer fidelity. Proving this result rests on the so-called *log-sum inequality* (Cover and Thomas, 1991):

$$\int p_1(x) \log \frac{p_1(x)}{p_0(x)} dx \geq \int p_1(x) dx \log \frac{\int p_1(x) dx}{\int p_0(x) dx}$$

with $p_0(x)$, $p_1(x)$ being probability functions. Equality occurs only when $p_0(x) = p_1(x)$. Note that this inequality demonstrates that the Kullback-Leibler distance is non-negative. To apply the log-sum inequality, note that the joint probability function for population's combined output is given by

$$p_Y(\mathbf{y}; \alpha) = \int \prod_{n=1}^N p_{Y_n|X}(y_n | x) p_X(x; \alpha) dx.$$

The KL distance between the noncooperative population's responses to two stimulus conditions α_0 and α_1 equals

$$D_{\mathbf{Y}^{(N)}}(\alpha_1 \parallel \alpha_0) = \int \left[\int \prod_{n=1}^N p_{Y_n|X}(y_n | x) p_X(x; \alpha_1) dx \right] \times \log \frac{\int \prod_{n=1}^N p_{Y_n|X}(y_n | x) p_X(x; \alpha_1) dx}{\int \prod_{n=1}^N p_{Y_n|X}(y_n | x) p_X(x; \alpha_0) dx} dy$$

Using the log-sum inequality with respect to the integral over the input X upper-bounds this distance and demonstrates the data processing theorem: $D_{\mathbf{Y}^{(N)}}(\alpha_1 \parallel \alpha_0) \leq D_X(\alpha_1 \parallel \alpha_0)$. Applying the log-sum inequality to the integral over Y_N , we find that $D_{\mathbf{Y}^{(N)}}(\alpha_1 \parallel \alpha_0) > D_{\mathbf{Y}^{(N-1)}}(\alpha_1 \parallel \alpha_0)$, with strict inequality arising as long as the individual neuron KL distances are not zero. Thus, as the population size increases, the KL distance strictly increases. Note that in this proof, homogeneity is not required: for *any* noncooperative population, homogeneous or not, the KL distance at the output increases with population size. Because the population distance cannot exceed the input distance, the output distance approaches some limit as the population grows. The question becomes what is the limit? A more detailed analysis (Sinanović and Johnson, 2004) shows that for *any* population, homogeneous or not, the population distance achieves its upper bound, making the information transfer ratio approach one asymptotically. Consequently, populations of a diverse or similar character can provide perfect information fidelity.

Acknowledgments

This work was supported by grants from the National Institute of Mental Health and from the National Science Foundation. The author benefited from discussions with R.M. Glantz and C.C. Lane.

Notes

1. We use the term “dependence” rather than “correlation” deliberately. Both terms mean that population outputs are not statistically independent of each other. Uncorrelated quantities can be statistically dependent, which makes correlation a weak form of statistical dependence.

2. A result known as de Bruijn's identity relates entropy and Fisher information (Cover and Thomas, 1991; Kang and Sompolinsky, 2001), but its utility is not obvious.
3. The correlation coefficient in the two-neuron Bernoulli model *does* capture all of the statistical dependence. If $\rho = 0$, the occurrence of spikes in the two neurons are statistically independent events. Note that some combinations of correlation coefficient ρ and spike probabilities p_1, p_2 are not permitted as they would yield negative probabilities for the population's probability function.
4. The example illustrated in Fig. 2 does exhibit synergy, but we could not determine whether a noncooperative population could yield the spike probability and correlation values in the synergy region.

References

- Abbott L, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11: 91–101.
- Alonso J-M, Usrey W, Reid R (1996) Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature* 383: 815–818.
- Berger T (1971) *Rate Distortion Theory*. Prentice-Hall, Englewood Cliffs, NJ.
- Bezzi M, Diamond M, Treves A (2001) Redundancy and synergy arising from correlations in large ensembles. *arXiv:cond-mat/0012119*.
- Cover T, Thomas J (1991) *Elements of Information Theory*. Wiley, New York.
- deCharms R, Merzenich M (1996) Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381: 610–613.
- Johnson D (2002) For top reasons why mutual information does not quantify neural information processing. In *Computational Neuroscience'02*, Chicago, IL.
- Johnson D, Gruner C, Baggerly K, Seshagiri C (2001) Information-theoretic analysis of neural coding. *J. Comp. Neurosci.* 10: 47–69.
- Johnson D, Gruner C, Glantz R (2000) Quantifying information transfer in spike generation. *Neurocomputing* 33: 1047–1054.
- Johnson D, Ray W (2004) Optimal stimulus coding by neural populations using rate codes. *J. Comp. Neurosci.*, to appear.
- Kang K, Sompolinsky H (2001) Mutual information of population codes and distance measures in probability space. *Phys. Rev. Letters* 86: 4958–4961.
- Kilgard M, Merzenich M (1999) Distributed representation of spectral and temporal information in rat primary auditory cortex. *Hearing Res.* 134: 16–28.
- Kolmogorov A (1956) On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Info. Th.* 3: 102–108.
- Nirenberg S, Carcieri S, Jacobs A, Latham P (2001) Retinal ganglion cells act largely as independent encoders. *Nature* 411: 698–701.
- Reich D, Mechler F, Victor J (2001) Independent and redundant information in nearby cortical neurons. *Science* 294: 2566–2568.
- Rozell C, Johnson D, Glantz R (2002) Information processing during transient responses in the crayfish visual system. In *Computational Neuroscience'02* Chicago, IL.
- Shamir M, Sompolinsky H (2001) Correlation codes in neuronal populations. In T. Dietterich, S. Becker, and Z. Ghahramani, eds.,

- Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA, Vol. 14, pp. 277–284.
- Shannon C (1948) A mathematical theory of communication. *Bell System Tech. J.* 379–423, 623–656.
- Sinanović S, Johnson D (2004) Toward a theory of information processing, in preparation.
- Strong S, Koberle R, de Ruyter van Steveninck R, Bialek W (1998) Entropy and information in neural spike trains. *Phys. Rev. Letters* 80: 197–200.
- Tishby N, Pereira F, Bialek W (1999) The information bottleneck method. In *Proc. 37th Allerton Conference on Communication, Control and Computing*, pp. 368–377.
- Warland D, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.* 78: 2336–2350.
- Wilke S, Eurich C (2002) Representational accuracy of stochastic neural populations. *Neural Comput.* 14: 155–189.