

Robustness and complexity co-constructed in multimodal signalling networks

Nihat Ay^{1,2}, Jessica Flack¹ and David C. Krakauer^{1,*}

¹Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

²Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22–26, D-04103 Leipzig, Germany

In animal communication, signals are frequently emitted using different channels (e.g. frequencies in a vocalization) and different modalities (e.g. gestures can accompany vocalizations). We explore two explanations that have been provided for multimodality: (i) selection for high information transfer through dedicated channels and (ii) increasing fault tolerance or robustness through multichannel signals. Robustness relates to an accurate decoding of a signal when parts of a signal are occluded. We show analytically in simple feed-forward neural networks that while a multichannel signal can solve the robustness problem, a multimodal signal does so more effectively because it can maximize the contribution made by each channel while minimizing the effects of exclusion. Multimodality refers to sets of channels where within each set information is highly correlated. We show that the robustness property ensures correlations among channels producing complex, associative networks as a by-product. We refer to this as the principle of *robust overdesign*. We discuss the biological implications of this for the evolution of combinatorial signalling systems; in particular, how robustness promotes enough redundancy to allow for a subsequent specialization of redundant components into novel signals.

Keywords: robustness; complexity; evolution; signal; information theory; multimodal

1. MULTIMODAL ETHOLOGY

There are many types of signals in animal communication. These signals can be classified according to five features: modality (the number of sensory systems involved in signal production); channels (the number of channels involved in each modality); components (the number of communicative units within modalities and channels); context (variation in signal meaning due to social or environmental factors); and combinatoriality (whether modalities, channels, components and/or contextual usage can be rearranged to create different meaning). In this paper, we focus on multichannel and multimodal signals, exploring how the capacity for multimodality could have arisen and whether it is likely to have been dependent on selection for increased information flow or on selection for signalling system robustness. The robustness hypothesis argues that multiple modalities ensure message delivery (backup signals; Johnstone 1996) when one modality is occluded by noise in the environment (Hauser 1997) or noise in the perceptual system of the receiver (Rowe 1999).

Some multimodal signals will have non-redundant features in which each modality elicits a different response. A compound stimulus can either elicit responses to (i) both components (OR function), (ii) only one of the two original components (XOR), (iii) a modulated version of the response to one of the two original components, or (iv) the emergence of an entirely new response (Partan & Marler 1999;

Flack & de Waal in press). For example, male jumping spiders (*Habronattus dosseus*) appear to communicate quality through the coordination of seismic and visual displays (Elias *et al.* 2003). In contrast to unimodal, multichannel signals, multimodal signals, like those used by the jumping spider, are not typically perceived as a single stimulus (Hillis *et al.* 2002). Receiver discrimination makes sense when the information content in each modality is not perfectly correlated (multiple messages; Johnstone 1996). Can multimodal signals in which modalities are not redundant also be explained by the robustness hypothesis, or is an alternative explanation required, for example, whether certain kinds of messages can only be communicated using compound stimuli?

We consider a simple model in which a signaller transmits a message to a receiver. Signaller and receiver are assumed to have matching interests and there is no advantage to deception. The signaller transmits the message through an arbitrary number of channels using, for example, multiple frequencies (e.g. the fundamental frequency, second harmonic, etc.) in a vocalization. The receiver is free to attend to as few or as many channels as it wishes. The signalling strategy is to generate correlations among the channels in such a way so as to allow the receiver to decode the intended meaning. We ask, how many channels and what correlational structure among the channels should the signaller use to allow the receiver to decode a message, assuming random subsets of channels become occluded?

The robustness of the message depends on two factors: the causal contribution of signalling channels to a receiver and the exclusion dependence on a channel following channel elimination. The causal contribution refers to the unique information provided

* Author for correspondence (krakauer@santafe.edu).

One contribution of 15 to a Theme Issue 'The use of artificial neural networks to study perception in animals'.

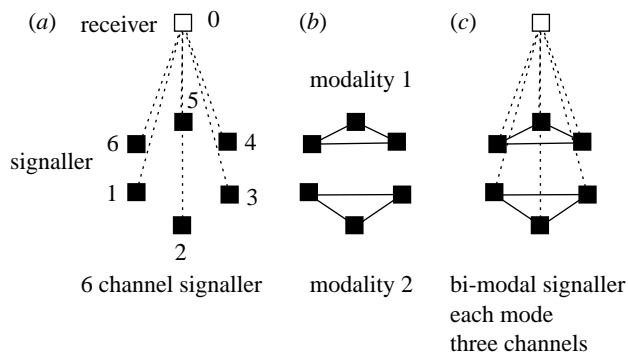


Figure 1. (a) The basic perceptron architecture illustrated with six nodes of a signaller (numbered black squares) and six corresponding channels used to convey a message generated by node activity to a receiver (white square numbered 0). (b) Patterns of correlated activity illustrated through connections among signaller nodes. Two sets of three nodes show highly correlated activity: nodes 1–3 are highly correlated and nodes 4–6 are highly correlated. Each correlated cluster is referred to as a modality. (c) Representation of the receiver integrating inputs from the six channels constituting two signalling modalities.

by each channel. As channels are duplicated, any one channel necessarily makes a smaller contribution to message meaning. Exclusion dependence refers to the consequences for decoding the message of occluding a single or a set of channels. To understand how this distinction maps onto communication in the natural world, consider the following vocalization example: experimental studies of the combination long calls of golden lion tamarins (*Sanguinus oedipus*) indicate that tamarins treat unmanipulated long calls as perceptually equivalent to long calls with deleted fundamental frequencies or second harmonics (Weiss & Hauser 2002). This illustrates the principle of causal contribution in which the absence of a single harmonic does not affect whether or how well a receiver decodes a call. However, tamarins do distinguish between unmanipulated calls and synthetic calls in which all of the harmonics above the fundamental have been deleted, or in which the second harmonic has been mistuned. This illustrates the principle of exclusion dependence in which removing the sets of channels, but not a single channel, can jeopardize accurate receiver decoding.

We represent the multichannel and multimodal property of signals through two different types of connectivity in a modified perceptron network. Channels connect the receiver to the nodes of a signaller (figure 1a), whereas clusters of highly correlated nodes of the signaller define modalities (figure 1b). In this paper, we show mathematically that the clusters of correlated activity constituting modalities emerge as robust solutions to a channel occlusion problem. We also find that our robustness measure acts as a lower bound on a well-known network complexity measure arising from maximizing information flow (Tononi et al. 1994). We end by discussing the potential implications for the evolution of combinatorial signals.

2. THE PERCEPTRON NETWORK OF RECEIVING

To describe the response of a receiver to the incoming signals, we consider a simple network structure in

terms of a map T , which describes how a node labelled 0 generates an output y based on the information from an input vector x_1, \dots, x_N (figure 1a). Let us assume \mathcal{A} for the set $\{1, \dots, N\}$ of input units and \mathcal{A}_0 for the set $\mathcal{A} \cup \{0\}$ of all units. The states of a unit $i \in \mathcal{A}_0$ are denoted by \mathbf{X}_i . The formal description of the transformation T is given by a Markov transition matrix

$$T : \mathbf{X}_{\mathcal{A}} \times \mathbf{X}_0 \rightarrow [0, 1], \quad (x, y) \mapsto T(y|x),$$

where T is the function performed by the network and the input set is given by $\mathbf{X}_{\mathcal{A}} := \mathbf{X}_1 \times \dots \times \mathbf{X}_N$. The value $T(y|x)$ is the conditional probability of generating the output y given the input $x = (x_1, \dots, x_N)$. This implies that for every $x \in \mathbf{X}_{\mathcal{A}}$,

$$\sum_{y \in \mathbf{X}_0} T(y|x) = 1. \quad (2.1)$$

Example. Nodes have two states ‘0 = not active’ and ‘1 = active’ corresponding to the presence or absence of an active input. The system parameters are the *edge/channel weights* w_i , $i \in \mathcal{A}$, which describe the strength of interaction among the individual input nodes i and the output node 0, and a *threshold value* θ for the output node which controls its sensitivity to the input. In neuroscience, the weight describes the product of the density of postsynaptic receptors and neurotransmitter, and the threshold controls the sensitivity. We assume that given an input vector $x = (x_i)_{i \in \mathcal{A}} \in \{0, 1\}^N$ in the first step, the output node assumes a value given by the function

$$h(x) = \sum_{i \in \mathcal{A}} w_i x_i - \theta,$$

and then, in the second step, it generates the output 1 with the probability

$$T^\beta(1|x) = \frac{1}{1 + e^{-\beta h(x)}}.$$

The normalization property (2.1) then implies that the output 0 is generated with probability $1 - T(1|x)$. Here, the *inverse temperature* β controls the stochasticity of the map T^β . This is the familiar perceptron neural network (McCulloch & Pitts 1943).

2.1 Information measures

In this section, we introduce a number of measures of information that form the basis of a formal definition of robustness in networks. The argument proceeds by relating the perceptron architecture, interpreted as a simple stochastic map, to Shannon information or Shannon entropy. The connection of entropy to information derives from their common roots in deriving an extensive measure of the degree of ignorance we possess about the state of a coarse-grained system. The greater our ignorance, the greater the information value of a signal. Stated differently, when we obtain information about a low entropy or highly regular system, we learn very little new.

Given an arbitrary subset $S \subseteq \mathcal{A}_0$, we write \mathbf{X}_S rather than the more cumbersome notation $\prod_{i \in S} \mathbf{X}_i$, and we consider the projection

$$\mathbf{X}_S : \mathbf{X}_{\mathcal{A}_0} \rightarrow \mathbf{X}_S, \quad x = (x_i)_{i \in \mathcal{A}_0} \mapsto x_S = (x_i)_{i \in S}.$$

With an input distribution p on \mathbf{X}_A and a stochastic map T from \mathbf{X}_A to \mathbf{X}_0 , we have the joint probability vector

$$\mathbf{P}(x, y) = p(x)T(y|x), \quad x \in \mathbf{X}_A, y \in \mathbf{X}_0. \quad (2.2)$$

The projection X_S becomes a random variable with respect to \mathbf{P} . Now consider the three subsets $A, B, C \subseteq \mathcal{A}_0$. The *entropy* of X_C or Shannon information is then defined as

$$H_{\mathbf{P}}(X_C) = - \sum_{z \in \mathbf{X}_C} \mathbf{P}\{X_C = z\} \ln(\mathbf{P}\{X_C = z\}).$$

This quantity is a measure of the uncertainty that one has about the outcome of X_C (Cover & Thomas 2001). Once we know the outcome, this uncertainty is then reduced to zero. This justifies the interpretation of $H_{\mathbf{P}}(X_C)$ as the *information gain* after knowing the outcome of X_C . Now, having information about the outcome of the second variable X_B reduces the uncertainty about X_C . More precisely, the *conditional entropy* of X_C given X_B is defined as

$$H_{\mathbf{P}}(X_C|X_B) = - \sum_{y \in \mathbf{X}_B, z \in \mathbf{X}_C} \mathbf{P}\{X_B = y, X_C = z\} \times \ln(\mathbf{P}\{X_C = z|X_B = y\}),$$

and we have $H_{\mathbf{P}}(X_C) \geq H_{\mathbf{P}}(X_C|X_B)$. Using these entropy terms, the *mutual information* of X_C and X_B is then defined as the uncertainty of X_C minus the uncertainty of X_C given X_B ,

$$I_{\mathbf{P}}(X_C : X_B) = H_{\mathbf{P}}(X_C) - H_{\mathbf{P}}(X_C|X_B).$$

The *conditional mutual information* of X_C and X_B given X_A is defined in a similar way,

$$I_{\mathbf{P}}(X_C : X_B|X_A) = H_{\mathbf{P}}(X_C|X_A) - H_{\mathbf{P}}(X_C|X_A, X_B).$$

We simplify the notation by writing these quantities without explicitly mentioning \mathbf{P} .

Thus, we have a measure of the information that is output by the network as a function of the information present at the input units.

3. NETWORK COMPLEXITY MEASURES AND SIGNALLING

Now that we have defined a simple signalling network and appropriate information measures, we discuss a measure of network complexity. This measure will refer to the structure of correlations among the nodes of the signaller (figure 1b) and lead to a statistical definition of a signalling modality. In a series of papers, Tononi et al. (1994; TSE) consider information theoretic measures of complexity in neural networks. The primary goal of this research is to determine which anatomical properties we should expect to observe in networks, such as a nervous system, where communication among cells plays a crucial role in promoting functional states of the system. TSE relate the functional connectivity of the network to statistical dependencies among neurons that arise through patterns of connectivity. The dependencies are measured using information theoretic expressions outlined in §2. TSE identify two principles of functional organization. The first principle derives from the observation that groups of neurons are

functionally segregated from one another; into modules, areas or columns. The second principle maintains that to achieve global coherence, segregated components need to become integrated. Segregation and integration combine to produce systems capable of both discrimination and generalization. In an animal signalling context, segregation can be related to clusters of cells dedicated to generating different messages, in other words, different modalities of expression. Integration binds these signals into a compound meaning or function. According to TSE, integration is a measure of the difference between the entropy expected on the basis of network connectivity and the observed entropy,

$$I(X_A) = \sum_{v \in A} H(X_v) - H(X_A).$$

The *TSE complexity* is then defined as

$$C(X_A) := \sum_{k=1}^N \left[\frac{k}{N} I(X_A) - \frac{1}{\binom{N}{k}} \sum_{\substack{A \subseteq \mathcal{A} \\ |A|=k}} I(X_A) \right]. \quad (3.1)$$

It has been shown that this complexity measure is low for systems whose components are characterized either by total independence or total dependence. It is high for systems whose components show simultaneous evidence of independence in small subsets and increasing dependence in subsets of increasing size. The TSE complexity can be written in terms of mutual informations,

$$C(X_A) := \sum_{A \subseteq \mathcal{A}} \frac{|A|}{N \binom{N}{|A|}} I(X_A : X_{\mathcal{A} \setminus A}).$$

Now that we have defined TSE complexity in closed form in terms of information measures, we can relate this back to information flows through the perceptron architecture.

4. ROBUSTNESS AS A GENERATOR OF COMPLEXITY

(a) A definition of robustness

In this section, we relate the TSE complexity to a robustness measure. In order to capture the main ideas behind this approach, we consider two random input variables X and Y with distribution $p(x, y)$, where $x \in \mathbf{X}$ and $y \in \mathbf{Y}$, and one output variable Z which is generated by a stochastic map

$$T : (\mathbf{X} \times \mathbf{Y}) \times \mathbf{Z} \rightarrow [0, 1], \quad (x, y, z) \mapsto T(z|x, y).$$

Now we assume that Y is knocked out, and we want to have a measure for the robustness of T against this knockout. First, robustness should include some notion of invariance with respect to this knockout perturbation. When the invariance property is satisfied, we say that the *exclusion dependence* is low. On the other hand, trivially vanishing exclusion dependence can be achieved if T does not at all depend on Y . In order to have a robustness measure that captures non-trivial invariance properties, we have to take the *contribution* of

Y to the function T into account, which can be done by applying Pearl’s formalism of causality (Pearl 2000). Our robustness measure is then defined as follows.

Informal definition of robustness. We define the robustness of T against knockout of Y as the contribution of Y to the function T minus the exclusion dependence of T with respect to the knockout of Y .

We consider the case where channels are occluded rather than simply noisy as limiting case that maps more naturally onto the biological problem that we are considering; namely, under conditions where channels are not available for inspection, how might alternative channels be used to extract the required information?

We (Ay & Krakauer in press) have formalized this probabilistic notion of robustness in terms of information geometry (Amari 1985) and derive the following formula:

$$R(Y, p, T) := \sum_x \sum_z \sum_y p(x, y) T(z|x, y) \ln \frac{\sum_{y'} p(x, y') T(z|x, y')}{p(x) \sum_{y'} p(y') T(z|x, y')}$$

where we sum over all values of x , y and z . This measures the amount of statistical dependence between X and Y that is used for computing Z in order to compensate the exclusion of y . The robustness vanishes if for all x and z ,

$$\sum_y p(x, y) T(z|x, y) = p(x) \sum_y p(y) T(z|x, y),$$

or equivalently

$$\sum_y T(z|x, y) (p(x, y) - p(x)p(y)) = 0. \tag{4.1}$$

There are two extreme cases where this equality holds. The first case is when there is no statistical dependence between x and y that can be used for compensation. Then, $p(x, y) = p(x)p(y)$ and the equality (4.1) holds. The other extreme case is when there is statistical dependence, but this dependence is not used by T . In this case, $T(z|x, y) = T(z|x, y')$ for all y and y' .

The ability of the perceptron to make use of redundant information by integrating over input channels is similar to von Neumann’s theory for probabilistic logics (von Neumann 1956).

(b) An example: duplication and robustness

In this section, let $T: \mathbf{X} \times \mathbf{Y} \rightarrow [0, 1]$ be a stochastic map and p be a probability distribution on \mathbf{X} . In this example, we seek to measure robustness as we duplicate T . In order to have several copies of this map, we consider the N -fold Cartesian product \mathbf{X}^N , and we define the input probability distribution

$$\tilde{p} := \sum_{x \in \mathbf{X}} p(x) \delta_{(x, \dots, x)}.$$

We define the extension of the map T to the set of N identical inputs by choosing one input node with probability $1/N$ and then applying the map T to that

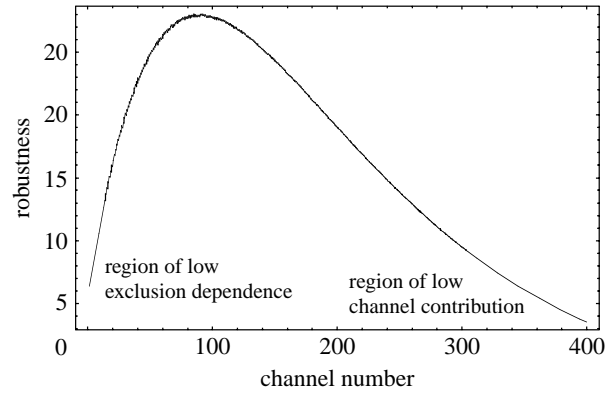


Figure 2. The robustness value as a function of channel number. At low levels of duplication, individual channels increase the robustness by lowering the exclusion dependence. At high levels of duplication, individual channels make very low contribution to function and thereby lower the robustness value. The figure illustrates that systems of large non-integrated elements should not be deemed robust as channel removal does not influence behaviour. For increasing numbers of channels to increase robustness, we need more than duplication, we require the emergence of correlated modules or statistical modalities. Parameter $\alpha = 10^{-3}$.

node. This leads to

$$\tilde{T}(y|x_1, \dots, x_N) := \frac{1}{N} (T(y|x_1) + T(y|x_2) + \dots + T(y|x_N)).$$

With the probability $1 - \alpha$ for the exclusion of an input node $v \in \{1, \dots, N\}$, we define the probability for a subset $A \subseteq \{1, \dots, N\}$ to remain as input node set after knockout as

$$r(A) := \alpha^{|A|} (1 - \alpha)^{N - |A|}.$$

We find the mean robustness of \tilde{T} with respect to this knockout distribution

$$R(r, \tilde{T}) := \sum_{A \subseteq \{1, \dots, N\}} r(A) \left[\begin{array}{l} \text{robustness of } \tilde{T} \text{ against} \\ \text{knockout of the} \\ \text{complement of } A \end{array} \right].$$

Now, we want to show the robustness properties with respect to the number N of channels and the probability α by specifying T as the identity map on the set $\{\pm 1\}$ with uniform distribution $p(-1) = p(+1) = 1/2$. The output node just copies the input: $x \mapsto x$. Following our concept of robustness, we can show that the robustness is given by

$$-\sum_{k=0}^N \binom{N}{k} \alpha^k (1 - \alpha)^{N-k} \ln \left(\frac{N+k}{2N} \right) - (1 - \alpha)^N \ln(2)$$

Figure 2 shows that channel duplication first leads to an increase of robustness but then declines as the number N increases. Note that the duplication of identical inputs is not optimal for robustness as each input transmits identical information and thereby lessens its contribution to the signal. Robustness would increase if each input overlapped but included some uncorrelated information.

(c) Extension to networks

In order to formally connect robustness with complexity, we extend the robustness measure to the network setting. The set of network nodes is denoted by V and the set of edges is denoted by $E \subseteq V \times V$. Given a unit $v \in V$, $\text{pa}(v) := \{u \in V : (u, v) \in E\}$ is the set of units that provide direct information to v . With the state sets \mathbf{X}_v , $v \in V$, we consider a family of stochastic maps denoted as

$$T^v : \mathbf{X}_{\text{pa}(v)} \times \mathbf{X}_v \rightarrow [0, 1], \quad (x, y) \mapsto T^v(y|x).$$

The global dynamics $T : \mathbf{X}_V \times \mathbf{X}_V \rightarrow [0, 1]$ is then defined by

$$T((y_v)_{v \in V} | (x_v)_{v \in V}) := \prod_{v \in V} T^v(y_v | x_v).$$

Now, we consider exclusions of subsets of the set V and associate a robustness measure for the network with regard to these exclusions. After knockout, we have a remaining set A , and for a node $v \in A$ we consider the part of $\text{pa}(v)$ that is contained in A and the part $\text{pa}(v) \setminus A$ that has been knocked out. Then, we have the following robustness of T^v against this exclusion:

$$\begin{aligned} & \text{Robustness of } T^v \text{ against exclusion of } \text{pa}(v) \setminus A \\ &= \sum_{x \in \mathbf{X}_{\text{pa}(v) \cap A}, y \in \mathbf{X}_{\text{pa}(v) \setminus A}, z \in \mathbf{X}_v} p(x, y) T^v(z|x, y) \\ & \quad \times \ln \left(\frac{\sum_{y' \in \mathbf{X}_{\text{pa}(v) \cap A}} p(x, y') T^v(z|x, y')}{p(x) \sum_{y' \in \mathbf{X}_{\text{pa}(v) \setminus A}} p(y') T^v(z|x, y')} \right). \end{aligned}$$

With a probability distribution r , we define the following total robustness of the network:

$$R(r, p, T) = \sum_{A \subseteq V} r(A) \left\{ \frac{1}{|A|} \sum_{v \in A} \left[\begin{array}{l} \text{Robustness of } v \\ \text{against exclusion} \\ \text{of } \text{pa}(v) \setminus A \end{array} \right] \right\}.$$

In this formula, we assume that p is a stationary distribution of T . Note that our robustness measure is a temporal quantity. It is surprising that one can relate this quantity to a purely spatial quantity, which depends only on the stationary distribution p . More precisely, we have the following upper bound for the robustness:

$$R(r, p, T) \leq \sum_{A \subseteq V} r(A) I(X_A : X_{V \setminus A}). \quad (4.2)$$

Let us compare this upper bound with the TSE complexity (3.1). For appropriate coefficients

$$r(A) = \frac{2|A|}{N(N+1) \binom{N}{|A|}}, \quad (4.3)$$

we have the following connection:

$$\sum_{A \subseteq V} r(A) I(X_A : X_{V \setminus A}) = \frac{2}{N+1} C(X).$$

Here, the coefficients are normalized, i.e. $\sum_{A \subseteq V} r(A) = 1$. With (4.2), this directly implies

$$R(r, p, T) \leq \frac{2}{N+1} C(X). \quad (4.4)$$

We see that for the special distribution (4.3), systems with high robustness display a high value of TSE complexity. To make the more general statement, we generalize TSE complexity to cases where we can arbitrarily select r . We assume

$$\begin{aligned} C(r, X) &:= \sum_{A \subseteq V} r(A) \left(I(X) - \frac{N}{|A|} I(X_A) \right) \\ &= \sum_{A \subseteq V} r(A) I(X_A : X_{V \setminus A}). \end{aligned}$$

In order to see how the above equation extends the definition (3.1), we consider distributions r with $r(A) = r(B)$ whenever $|A| = |B|$. Such a distribution is uniquely defined by a map $r : \{0, 1, \dots, N\} \rightarrow [0, 1]$ with $\sum_{k=0}^N \binom{N}{k} r(k) = 1$. This implies

$$C(r, X) = \sum_{k=0}^N \frac{r(k)N}{k} \binom{N}{k} \left[\frac{k}{N} I(X_A) - \frac{1}{\binom{N}{k}} \sum_{\substack{A \subseteq V \\ |A|=k}} I(X_A) \right].$$

We find that this is closely related to the complexity definition (3.1). The only difference is that in the generalized definition the sum is weighted appropriately. In any case, we have an extension of the inequality (4.4),

$$R(r, p, T) \leq C(r, X). \quad (4.5)$$

This equation makes explicit that network robustness is a lower bound on network complexity. Hence, pressure towards an increase in robustness, through for example multimodality, will always lead to an increase in the complexity of the message by promoting structured correlations in the channels. TSE have found through extensive numerical simulation that the complexity measure $C(X)$ is maximized by network structures in which cells form densely connected clusters which are themselves sparsely connected, in other words, partially overlapping functional modules. In the context of signalling networks, segregated modules can be interpreted as different signalling modalities, where each modality emphasizes a different feature of the system. Of interest to us in this paper is the way in which signal complexity arises naturally out of pressures favouring robust signal detection.

5. MULTIMODAL ROBUSTNESS AND COMBINATORIAL COMPLEXITY

We have derived a measure to quantify the impact of signalling channel perturbations on the ability of a receiver to process a signal. The measure quantifies a functional distance between a ‘perfect information’ condition between signaller and receiver and a reconfigured condition in which a subset of input channels have been occluded. Using multiple channels through which correlated activity can be transmitted by a signaller allows a receiver to choose from a different channel once one has been lost. However, if all channels are identical, then it makes little sense to refer to a signalling system as robust following removal. This is because as channels become more numerous, each channel makes a diminishing causal contribution

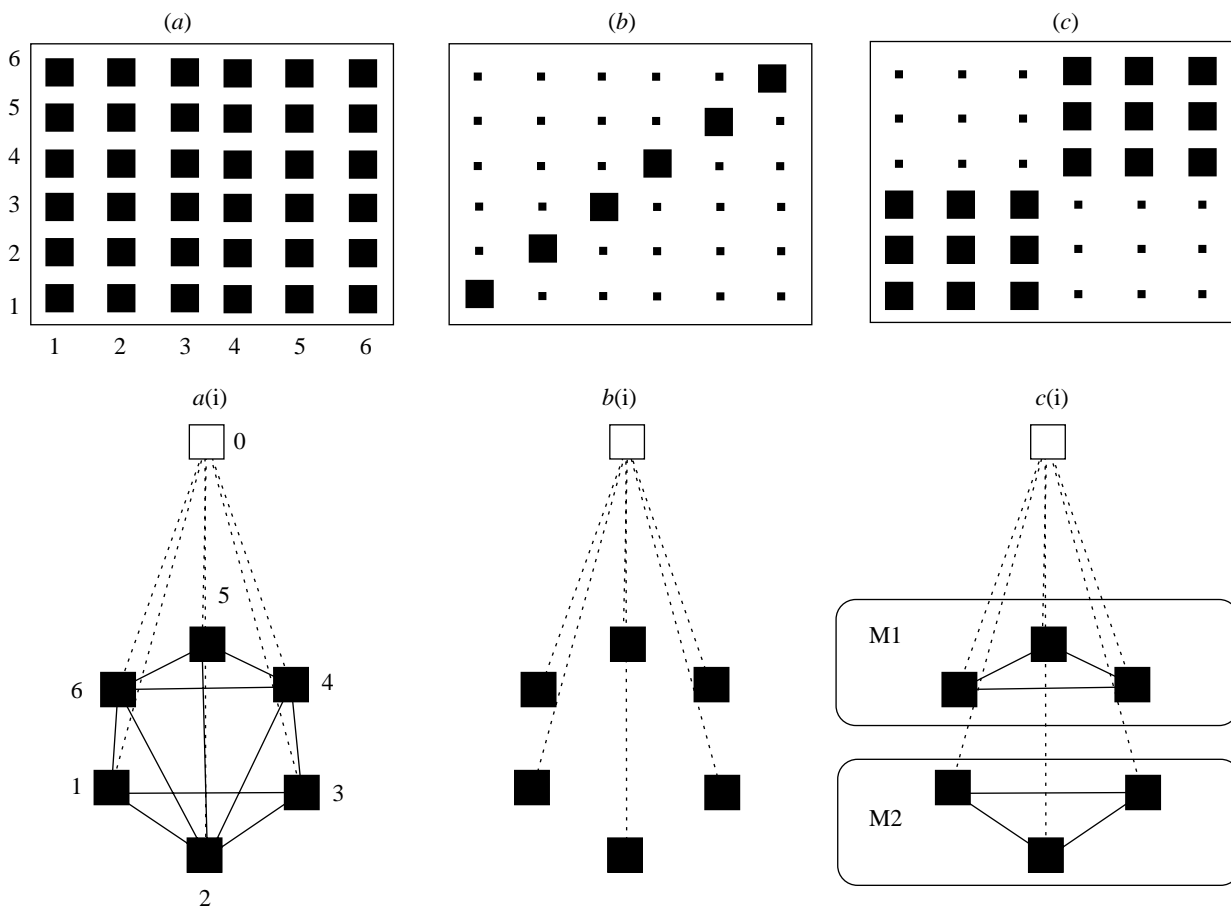


Figure 3. (a) Correlation matrix indicating by the size of black squares the magnitude of correlation or mutual information between pairs of nodes of the signaller. Here, all nodes are perfectly correlated in their activity. (a(i)) The perceptron connectivity corresponding to the correlation matrix (a). (b) Signaller nodes are only weakly correlated with each other and constitute approximately independent channels (b(i)) for the receiver to integrate. (c) Channels form correlated clusters of activity, with weak correlations among clusters. This corresponds to a two-modality signalling system (sets of channels M1 and M2). The two-modality case is both more robust and more complex. Robustness derives from insensitivity to channel occlusion through channel redundancy coupled to high information flow through weak modality decoupling.

to the message (Pearl 2000). In the large channel limit, each channel becomes effectively independent of the message (figure 3a) as its unitary contribution is just $1/N$. From the selection perspective, irrelevant channels are expected to be lost (Krakauer & Nowak 1999). By reducing correlations among channels, we increase their individual causal contribution to the message. However, if they become completely decorrelated (figure 3b), the removal of any one has a significant impact on the receiver's ability to decode the message. The optimal solution is to generate clusters within which channel activity is highly correlated and between which activity is weakly correlated (figure 3c). Hence, an optimal signaller distributes a message among weakly correlated modalities within which multichannel redundancy remains high, each cluster we might think of as a primordial modality promoting specialized interpretive means by the receiver.

Somewhat surprisingly, the modular structure of a robust signal is what makes for an effective information processing neural network (Tononi *et al.* 1994). Selection for high information flow and thereby reduced sensitivity to channel occlusion leads naturally to signalling networks with high levels of integration

and segregation (clustering). High levels of segregation and integration in turn promote the development of an effective computational system (in the signaller) which trades off specificity with generalization. This property has been referred to as network complexity (Tononi *et al.* 1994).

In terms of real communication systems, our results suggest that multimodality might arise in the following way: senders emit multichannel signals (like the tamarin combination long call or chimpanzee pant grunts, each of which is characterized by the presence of multiple frequencies). Over time, selection for robustness generates correlations among channels and, consequently, channel clustering. From the signal production/encoding perspective, this means that neural and behavioural substrates underlying production are becoming modularized, setting the stage for the evolution of alternative sensory modalities. Pant grunts, a chimpanzee subordination signal, are no longer produced alone, but are emitted in conjunction with gestural or behavioural displays, like bobbing and bowing (de Waal 1982). From the decoding perspective, this means that the cognitive cost of being able to perceive and process multiple redundant channels is minimized in so far as the

overlap between modalities (clusters of channels) is just great enough that each mode contributes to signal meaning and yet is sufficient to decode signal meaning if the other mode is knocked out.

It is worth saying a few more words about the relationship between complexity and robustness. In the complexity measure, random bi-partitions of a network (figure 1b) are used to assess the extent of communication among network regions of a signaller; whereas in the robustness measure, bi-partitions reflect removals of large sets of communicating channels. In both cases, information flow is required. In the complexity case, information flow is assumed to reflect increased associative power among modules; whereas in the robustness case, information flow is required in order that non-occluded channels can be used as alternative information sources. It is not obvious why selection should favour information flow among channels of a signaller, but it is obvious that this information flow can be used by the receiver in case of occlusion. If we were considering a signal internal to the sender, the situation would be different as in this case information flow could come under selection for more effective integration for cognitive function.

The net result we found is that signallers seeking to transmit complex messages benefit from a multimodal strategy as it both increases the diversity of information flowing to a receiver and increases the robustness of the signal. That robustness leads to complexity through a shared requirement for increased information flow has been termed the *principle of robust overdesign* (Krakauer & Plotkin 2004).

A secondary advantage of selection for increased information flow is that it provides the basis for combinatorial signalling. Signals can be built up combinatorially out of segregated functional units, with each unit producing a different signal component. Integration ensures that at first these components have overlapping meanings (through correlated activity) and are thereby likely to be understood or learnable. This applies to multicomponent signals (different communicative features in the same modality) as well as to multimodal ones. The learnability problem (Flack & de Waal *in press*), which is particularly problematic when receivers are confronted with new signals that are spatially or temporally divorced from their objects and thereby hard to associate, can be mitigated by 'pointing' to a new signal object using a 'compound stimulus'. A compound stimulus signal is one comprising two or more modes or components, each of which has an established meaning. Emitting these together allows the receiver to infer from overlap a new meaning. The capacity for combinatoriality is additionally advantageous because many meanings can be created from a small set of components, reducing the need for cognitively burdensome and error-prone storage of many one-to-one mappings—for example, it is easier to generalize unknown word meaning from contextual usage (Grice 1969) than it is to store every word (and their associated meanings) one is likely to encounter (Nowak & Krakauer 1999). Thus, with the evolution of signal robustness, there is the possibility of piggybacking a complex, integrated signalling system with the potential for combinatoriality and increased message encoding.

Selection for perfect combinatoriality (independent channels or modalities) would presumably favour decreasing overlap in meaning among signal components, thereby decreasing robustness in the long run. An interesting open question is what level of component divergence is optimal, given selection for combinatoriality in the context of robustness constraints.

REFERENCES

- Amari, S. 1985 *Differential-geometric methods in statistics. Lecture notes in statistics*, vol. 28. Heidelberg, Germany: Springer.
- Ay, N. & Krakauer, D. C. *In press*. Geometry robustness theory and biological networks. *Theory Biosci.*
- Cover, T. M. & Thomas, J. A. 2001 *Elements of information theory*. New York, NY: Wiley.
- de Waal, F. B. M. 1982 *Chimpanzee politics: power and sex among the apes*. Baltimore, MD: Johns Hopkins Press.
- Elias, D. O., Mason, A. C., Maddison, W. P. & Hoy, R. R. 2003 Seismic signals in a courting male jumping spider (Araneae: Salticidae). *J. Exp. Biol.* **206**, 4029–4039.
- Flack, J. C. & de Waal, F. B. M. *In press*. Context modulates signal meaning in primate communication. *Proc. Natl Acad. Sci. USA*.
- Grice, H. P. 1969 Utterer's meaning and intention. *Phil. Rev.* **68**, 147–177. (doi:10.2307/2184179)
- Hauser, M. D. 1997 *The evolution of communication*. Cambridge, MA: MIT Press.
- Hillis, J. M., Ernst, M. O., Banks, M. S. & Landy, M. S. 2002 Combining sensory information: mandatory fusion within, but not between senses. *Science* **298**, 1627–1630. (doi:10.1126/science.1075396)
- Johnstone, R. A. 1996 Multiple displays in animal communication: 'backup signals' and 'multiple messages'. *Phil. Trans. R. Soc. B* **351**, 329–338.
- Krakauer, D. C. & Nowak, M. A. 1999 Evolutionary preservation of redundant duplicated genes. *Semin. Cell Dev. Biol.* **10**, 555–559. (doi:10.1006/scdb.1999.0337)
- Krakauer, D. C. & Plotkin, J. B. 2004 Principles and parameters of molecular robustness. In *Robust design: a repertoire for biology, ecology and engineering* (ed. E. Jen), pp. 71–103. Oxford, UK: Oxford University Press.
- McCulloch, W. S. & Pitts, W. A. 1943 Logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133. (doi:10.1007/BF02478259)
- Nowak, M. A. & Krakauer, D. C. 1999 The evolution of language. *Proc. Natl Acad. Sci. USA* **96**, 8028–8033. (doi:10.1073/pnas.96.14.8028)
- Partan, S. & Marler, P. 1999 Communication goes multimodal. *Science* **283**, 1272–1273. (doi:10.1126/science.283.5406.1272)
- Pearl, J. 2000 *Causality*. Cambridge, UK: Cambridge University Press.
- Rowe, C. 1999 Receiver psychology and the evolution of multicomponent signals. *Anim. Behav.* **58**, 921–931. (doi:10.1006/anbe.1999.1242)
- Tononi, G., Sporns, O. & Edelman, G. M. 1994 A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl Acad. Sci. USA* **91**, 5033–5037. (doi:10.1073/pnas.91.11.5033)
- von Neumann, J. 1956 Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Automata studies* (eds C. E. Shannon & J. McCarthy), pp. 43–98. Princeton, NJ: Princeton University Press.
- Weiss, D. J. & Hauser, M. D. 2002 Perception of harmonics in the combination long call of cottontop tamarins (*Saguinus oedipus*). *Anim. Behav.* **64**, 415–426. (doi:10.1006/anbe.2002.3083)