Information and the function of neurons

Marc Burock

A B S T R A C T

Many of us consider it uncontroversial that information processing is a natural function of the brain. Since functions in biology are only won through empirical investigation, there should be a significant body of unambiguous evidence that supports this functional claim. Before we can interpret the evidence, however, we must ask what it means for a biological system to process information. Although a concept of information is generally accepted in the neurosciences without critique, in other biological sciences applications of information, despite careful analysis, remain controversial. In this work I will review classical stimulus-response studies in neuroscience and use Claude Shannon's mathematical information theory as a starting point to interpret information processing as a function of the brain. I will illustrate a disanalogy between Shannon's communication model (source, encode, channel, receiver, decode) and neural systems, and will argue that the neural code is not very code-like in comparison to genetic and engineered codes. I suggest that we have conflated the act of representing neuroscientific facts—which we do to summarize and communicate our findings with others—with taking experimental facts to be representations.

Keywords: function; evidence; information theory; neuroscience; representation

1. Introduction

Many cognitive scientists and neuroscientists, and perhaps most people in general, believe that the brain processes information, although there is ambiguity about what this belief entails. Bechtel and Richardson (2010), as philosophers of cognitive science, consider it uncontroversial that cognitive scientists involved in neuroimaging research believe that "the brain contains some regions that are specialized for processing specific types of information" (p. 241). Neuroscientists too claim that "the principal function of the central nervous system is to represent and transform information" (deCharms & Zador 2000, p. 613). Given such wide-spread acceptance of a belief, it is appropriate to ask for the justification of this belief. If the justification is empirical and experimental, then we should look to the research reported by working scientists in the field; if it is primarily theoretical, then we should look to the arguments of philosophers and theoreticians.

We will no doubt discover both kinds of justification if we look for it. Yet we also assume that scientists, when stating that the brain processes information, are primarily stating an empirical fact or a widely agreed-upon scientific proposition that is supported by a body of experimental evidence. Like the physiologist who can back up the proposition 'kidneys filter the blood' with a presentation of the experimental evidence, we expect that the neuroscientist should be able to do the same regarding a functional claim about the brain. I am not suggesting a definition of science or attempting to solve Popper's demarcation problem, but am appealing to the belief that widely accepted scientific statements ought to be associated with unambiguous evidence.

The task here, however, is somewhat more involved than an objective review of the scientific literature. The concept of information processing, and informational language in general, has received considerable critique within the field of biology, with authors disagreeing about the explanatory and theoretical weight of informational concepts (Sarkar 1996; Maynard Smith 2000; Godfrey-Smith 2000; Griffiths 2001). It is, however, generally agreed that the concept of information developed by Claude Shannon as used in mathematical information theory can be appropriately applied to the biological sciences (Godfrey-Smith and Sterelny 2008). Shannon's concept of information can, in fact, be applied to *anything* that can be represented as a random variable (Cover & Thomas 2006). Given

the broad applicability of Shannon information, it is helpful to further define the concept of *carrying* Shannon information. Objects are said to carry Shannon information about each other when the states of two or more objects are physically or causally correlated (Dretske 1981; Piccinini & Scarantino 2011).

Scientists, however, in saying that the function of the brain is to process information must be saying more than this. Smoke carries Shannon information about fire, for the presence or absence of fire and smoke are causally correlated, but certainly information processing in the brain means something more than causal correlation. More, authorities claim that information processing is the *function* of the brain. Although smoke carries information about fire, most people would not claim that it is the function of smoke to do so. In this work I analyze a broad category of neuroscientific evidence to determine (1) if this evidence supports a richer concept of information processing than causal correlation, and (2) if this richer concept can unambiguously be called the function of the brain. I will argue that this evidence does not support a richer concept of information in the neurosciences, and that the functional attribution is either empirically unjustified or too ambiguous for careful theoretical use.

Cognitive scientific evidence, especially neuroimaging evidence, has been increasingly subjected to criticisms. To better demarcate my position, I highlight that I am not specifically arguing between distributed versus localized processing in the brain (Utel 2001; Hardcastle & Stewart 2002; Bunzl et al. 2010), or pointing out the previously discussed technical-methodological limitations of brain assessing technologies (Logothetis 2008; Roskies 2007; Klein 2009). I do share with these authors the broader concern for interpretations of evidence in the field of cognitive science, and how theoretical assumptions influence interpretations of evidence, ultimately ending in statements made by researchers that carry the weight of scientific fact. These facts, in turn, are used by naturalistic philosophers of mind to constrain philosophical theory and argument.

2. External-stimulus/brain-response studies and neuronal function

Although there are other categories of neuroscientific evidence, I will be focusing exclusively on external-stimulus/brain-response (ES/BR) studies as these experiments have been traditionally used to justify claims of information processing in the brain. In ES/BR studies the experimenter systematically manipulates physical features of an organism's external environment and measures temporally coincident properties of the organism's brain. The brain responses (BR) need not occur precisely simultaneous with the stimulus and are typically extended in time. Brain responses in ES/BR studies are recorded using a variety of techniques based upon electromagnetic brain properties, including single-unit intra and extracellular recording, evoked potentials, EEG, MEG, fMRI, and others. Edgar Adrian is generally credited with pioneering stimulus-response studies of nervous tissue. He was the first to record the electrical activity of single nerve fibers, the first to use the term information to describe neuronal activity (Garson 2003), and was subsequently awarded a Nobel Prize in 1932 for his work.

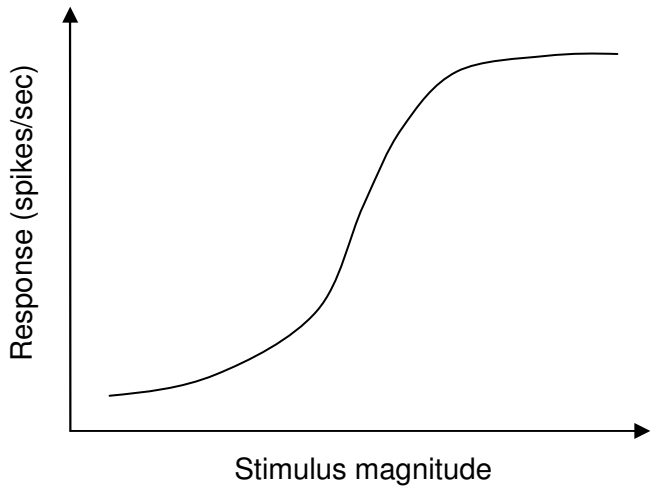

Fig 1. Schematic of a classic sigmoidal stimulus-response curve. The spike rate is a function of stimulus magnitude, allowing one to map spike rates onto stimulus magnitudes and vice-versa

2.1 Examination of the evidence

Adrian and Zotterman (1926), in their ground-breaking ES/BR research, measured the electrical responses of single sensory stretch receptors while they were fixed to varying weights. Adrian and Zotterman observed that a cell's electrical responses are in the form of stereotyped action potentials, or spikes, and that the rate of producing spikes increases as the weight increases. Thus the rate or frequency of spikes during a fixed time period is able to predict the

magnitude of the stimulus. Spike-rate can be plotted as a function of stimulus magnitude, demonstrating what is meant by a neural rate-code (Fig. 1).

These early experiments established that single cell responses and stimulus magnitudes may reliably covary with each other. While magnitudes and intensities are important properties of stimuli, they are not the only properties of environmental stimuli that are relevant to an organism. In general, a stimulus may be characterized by multiple properties. For example, an auditory stimulus may be described by its intensity, frequency spectrum, temporal envelope, source direction, source distance, and so on. It is possible that a particular cell responds to one of these properties and not to others, or to some combination of properties, which suggests that a cell may be *selective* for specific properties or features of the stimulus.

Barlow (1953) was perhaps the first to clearly demonstrate the feature selectivity of sensory cells (Reike et al. 1999). By recording the electrical activity of retinal ganglion cells in the frog, he was able to show that the cell's activity covaries with the location and size of a circular spot light on the retina. After systematically varying the light spot's size and location, Barlow determined that the cell's receptive field—the collection of stimulus properties that maximally activated the cell—is a circularly symmetric form called a center-surround field. Spots of light within a small region of the retina activate the cell, but spots of light away from that region inhibit it.

Hubel and Wiesel (1962) greatly extended Barlow's work and discovered cells of the striate (visual) cortex that have surprisingly complicated receptive fields. Two of these cell types are the so-called simple and complex cells, which respond maximally to appropriately oriented bars or slits of light. Some of the cells are relatively insensitive to the location of the bar, while others only appreciably respond to moving bars. In describing these cells, Hubel says that

> We feel that we have at least some understanding of a cell if we can say that its duty is to take care of a 1 degree by 1 degree region of retina, 6 degrees to the left of the fovea and 4 degrees above it, and to fire whenever a light line on a dark background appears, provided it is inclined at about 45 degrees. (Hubel 1962, p. 168)

The evidence from these pioneering ES/BR electrophysiological studies cannot be interpreted without the concept of selective response. Selective response means, loosely, that the cell fires action potentials only when the 'right' stimulus is present. Put more rigorously, selective response refers to two characteristics of neuronal cells: (1) the rate or *pattern* of firing action potentials (the spike train) covaries with specific stimulus properties, and (2) different cells may respond differently to the same stimulus. Both characteristics are typically implied when referring to the selectivity of cells in ES/BR studies. If someone discovered a neuron that exhibited (1), but on subsequent research discovered that all neurons exhibited (1) in the same way, one would not say that the initial neuron was selective for the stimulus, even though it exhibited selectivity for some stimuli among others. As well, the fact that different neurons respond differently to similar stimuli does not imply (1), since neuronal responses may be random in response to stimuli. Condition (1) is a form of *within* neuron stimulus selectivity, while condition (2) is a form of *between* neuron stimulus selectivity. For ES/BR studies such as Hubel and Wiesel's, when an ES is chosen and controlled by the researcher, we assume that the relation between the ES and BR is causal, as this assumption does not change our interpretation of selectivity, even though we use the term 'covaries' which has statistical connotations.

We are now in a position to evaluate whether Hubel and Wiesel's ground-breaking ES/BR studies justify the claim that the brain processes information in a way that means more than causal correlation. The experimental evidence consists of recorded responses of complex cells that demonstrate stimulus selectivity in the senses of (1) and (2). It seems that selectivity in the sense of (2) does not provide any justification that complex cells process information; the fact that different cells respond differently to the same stimulus suggests only that the cells are different in some way.

Claims of information processing, if they are justified by this experiment, must follow from the evidence that complex cell spike trains covary with the properties of

visual stimuli, or in causal language, that different visual stimuli cause different complex cell spike trains. Considering the latter causal language, the fact that different causes reliably produce different effects when mediated by the same cell does not appear to justify the claim that the cell processes information in a sense other than Shannon information. Even so, this type of causal relationship appears everywhere one looks. A particular pool ball when hit by other balls with different masses and velocities will undergo different effects. The pool ball may not appreciably move when stimulated by light or sound at typical intensities. The selectivity of the pool ball to acquire different velocities in response to different causal 'stimuli' does not appear fundamentally different than the selectivity of a complex cell, especially if the visual stimulus is taken to be a space-time collection of photons.

On closer analysis, there is a difference between the causality in the pool ball example and the relation between the ES and BR of complex cells. The pool ball example involves direct physical contact and an exchange of energy and momentum, while the causal response of the complex cell is more indirect. Photons travel through the lens of the eye and are absorbed by photoreceptor cells of the retina. Absorption of photons modulates the release of the neurotransmitter glutamate at synapses onto so-called bipolar cells, causing the electrical field across the membrane of these cells to become more positive or negative, which respectively increases or decreases the probability of generating an action potential. Bipolar cells have axons that synapse on other cells, and through a series of neuronal connections, influence the membrane potential of complex cells and subsequent action potential generation. The causal chain from photons to complex cell response is complicated and likely includes causal feedback, yet it is not obvious that a complicated causal chain is necessarily information processing.

Even more worrisome is the fact that selective causation need not imply that the BR has any *functional* relation to the ES at all. Nothing rules out the possibility that those selective correlations are accidental—not in the sense that the correlations are statistically spurious, but that those correlations are functionally irrelevant to the stimuli of interest. As an analogy, suppose my computer has a CPU fan with a blue LED light on the fan. The light, however, is unlit and the fan isn't spinning. It happens that when I kick my computer just so on the left side of the front cover, the LED lights up, the fan begins spinning but stops after a second or two, and the light goes out. If I kick it again, just so, it starts up for a second then stops. I can reliably cause the fan to turn on for a bit. When I kick the computer in other places, or shake it up, or sing to it, nothing happens to the fan. The fan is selectively correlated with a specific kick. Perhaps there are hundreds of computers, constructed at the same factory, that behave similarly. This selective, causal, complex, and perhaps arbitrary relationship does not imply that the fan is functionally relevant to my kicking, or processes kicking information, or represents kicking—the relationship may be accidental.

2.2 Functional claims

Intuitively, there appears to be a qualitative difference between the computer fan example and neurons like center-surround ganglion cells. The correlation between light-rays and the firing of ganglion cells is beneficial to the organism in some way, while the correlation between kicking the computer and the spinning fan is not beneficial to the computer—but this first hint at a difference is not convincing. If my computer fan is not spinning, and the CPU is rapidly heating up which may cause the computer to crash, then the correlation will be quite beneficial if I start kicking my computer.

We must look elsewhere to explain the difference between the two cases, and this search leads one to consider fundamental issues in biology concerning purpose, function, and design. The computer was not designed (by an engineering team) to manifest a correlation between kicking and fan movement, while the organism was designed (through natural selection or else wise) to have ganglion cells whose activity covaries with patterns of light rays. It is highly controversial whether artificial design and natural selection can be grouped into a univocal concept of design to support the above intuition, although some authors clearly make use of a broad notion of design to do so (Kitcher 1993).

We would like to say that it is the function of ganglion cells to produce activity that correlates with light rays, and it is not the function of the fan to correlate with kicking the computer—the latter is an

accident. Once we proffer this explanation, we must acknowledge that selective correlation, by itself, is not sufficient to establish the existence of a function, information processing or otherwise. This finding is no surprise to philosophers of biology who have investigated the concept of biological function over the past fifty years, but it reminds us of the limitations in arguing that it is the function of the brain (or neurons) to process information based upon the discovery of selective correlations between stimuli and neuronal activity.

Philosophers of biology have understood biological functions in numerous ways, but two formulations are prominent, one grossly characterized as 'backward-looking' and the other as 'forward-looking'. The backward looking or 'etiological' concept roughly defines a function of a given trait in terms of the trait's causal history of effects (Wright 1973; Millikan 1984; Neander 1991). It is often called the *selected effect* concept of function because it takes biological functions to be (historically) casual consequences that were preserved via natural selection. The forward looking concept, in contrast, defines a function of a trait in terms of its causal dispositions or capacities, where the relevant capacities are often taken to advance some goal or purpose of the organism (Rudwick 1964; Bigelow and Pargetter 1987), although Cummins (1975) proposed a dispositional theory without direct reference to purpose that is perhaps most acceptable to practicing scientists (Amundson and Lauder 1994). Roughly, Cummins' theory of function, often called a *causal role* theory, involves relating the causal capacities of parts of a system to other capacities of the whole system. A part-wise capacity that contributes to a capacity of the whole is said to be a function of that part.

When experimental neuroscientists like Hubel and Wiesel identify selective correlations between light patterns and ganglion activity, what additional observations or assumptions allow them to claim that the function of ganglion cells is to produce activity that correlates with the presence and absence of light patterns? Selective correlation is not enough to establish function, and it does not seem that experimental neuroscientists need appeal to natural selection in order to justify functional claims. Scientists identify or at least speculate on function by investigating the systems of interest in the lab, thereby explicating causal mechanisms that, for instance, make vision possible. We presume that the correlation between light patterns and ganglion activity contributes to the capacity of the organism to see—the correlation plays a causal role in vision—whereas the correlation between kicking and the spinning fan is not connected to any (relevant) capacity of the computer.

Correlation, however, is not the end of the story and is not a particularly useful identification of function. Many biological variables correlate, both within the organism and with the environment. Heart rate and the speed of ambulatory locomotion in humans covaries, but it would be a odd to say that it is the function of the heart to produce contraction rates that correlate with the speed of movement. When we attribute the function of information processing to neurons, we appear to be attributing something more than correlation or selective causal response. But to infer information processing from an experimentally observed ES/BR correlation, we must first assume that the correlation is not accidental in the sense above. A correlation can be said to be non-accidental if the organism was designed to manifest that correlation, if it was maintained by a selection process, or if the correlation is functional in the sense that the correlation contributes to some relevant capacity of the organism. The latter notion of a non-accidental correlation is perhaps the least controversial among practicing experimental neuroscientists, but if we use this understanding to support claims of information processing, then information processing, with respect to ganglion cells, simply means a correlation between the ES and BR that contributes to the capacity for vision.

The lack of richness in this functional claim, based upon the evidence, does not significantly depend upon a causal role theory of function. If one wishes to apply a selected effect notion of function to neuroscience data like in Garson (2011), then we can say that the brain structure was selected (by a neural selective process) to manifest a correlation between the ES and BR. Neither notion of function appears to transform the empirical correlation into a theoretically useful concept of information processing.

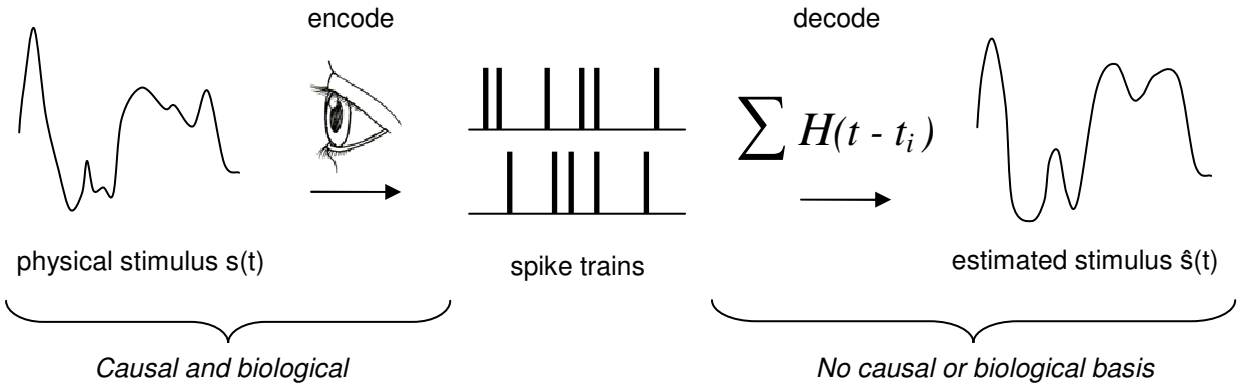

Fig 2. The typical encoding-decoding relation used in the neurosciences. Physical stimuli are encoded into spike trains through a causal, biological process. Spike trains are decoded into estimates of stimuli using mathematical heuristics that are unrelated to the biology of the organism

3. Justification of information processing from ES/BR studies

There is a strong tendency to associate information processing with the results of ES/BR experiments like Hubel-Weisel's. The spike trains of neurons appear to be relaying specific messages about the external environment to the organism. Claude Shannon (1948), the founder of mathematical communication theory, rigorously defined a model of information transfer that may explain this appearance. In Shannon's language, the physical environment acts as a source that generates a message (ES), the message is transformed by an encoder—a sensory organ of the organism—into a signal suitable for biological transmission. The spike train (BR) is assumed to be this signal and the neuron to be the transmission channel. These comparisons are reasonable, but the next stage of the communication model, however, is problematic (see Fig. 2). Communication requires a receiver that performs the inverse operation of the transmitter, or something that reconstructs the environmental message from the spike train signal.

The experimental researcher, the one who discovers selective correlations between neuronal spike trains and environmental messages (stimuli), often plays the *surrogate* role of the receiver or decoder. By describing relational or mathematical mappings between the ES and BR, neuroscientists attempt to 'read the neural code.' But this is not the sort of information transmission we were trying to explain. To complete the biological communication model, and to ground information transfer, we need to explain how the organism can reconstruct the environmental message from its temporal pattern of action potentials, and we must demonstrate that the organism reproduces a similar environmental message within the organism itself. The neuronal spike train is not the message—if anything it is the transmission signal or encoded message. Although interesting, it is not enough to show that spike trains have the *capacity* to represent environmental messages through selective covariation. The fact that researchers can mathematically map spike trains back onto stimuli does not say anything about how the organism biologically reconstructs the environmental message. This capacity to map follows immediately from statistical correlations. Neuroscientists who acknowledge these limitations explain that mathematically reconstructing stimuli from spike trains requires taking the homunculus point of view (Reike et al. 1999).

For an organism to receive an environmental message in Shannon's sense, that message must be within the organism and have the same structure as the original message. This suggestion may appear radical, but it is simply the completion of Shannon's communication model—the same model that supports the intuition that the brain processes and transmits information. For example, consider telephonic communication. Air pressure waves may be converted into analog electronic messages that are encoded into

digital signals and transmitted through a physical channel. This digital signal, which does not mirror the sound wave in form, reaches a destination where it is reconstructed back into an analog message that drives a loudspeaker, reproducing the original pressure wave. If the original message was not reproduced (perhaps imperfectly) at a destination, we could not claim that communication or information transfer took place. A message is communicated if and only if that message is reproduced at the receiver.

If one assumes that the organism receives environmental messages, then in accordance with Shannon's communication model, at least the structure of that message must be physically reproduced within the organism. The alleged *encoded* message—or spike train—has a physical basis, thus the message ought to have a physical basis as well. This means that the scientist would have to demonstrate a set of brain-related physical measurements that copy, perhaps imperfectly, the structure of an environmental stimulus. Let us call this the brain-image of an environmental message. It would remain for the scientist to describe the mechanisms by which neuronal spike trains causally reconstruct the brain-image of a particular environmental message.

When decoding spike trains in practice, the neuroscientist leaves the animal lab and goes to work at the computer. On the computer, spike trains and environmental stimuli are given numerical representations. The creative work involves finding mathematical algorithms and heuristics—let us call these the decoding procedures—that link spike trains to stimuli. When the neuroscientist finds a decoding procedure that works, she claims to have discovered a neural code. The problem is that the neurons themselves have no physical relation to the decoding procedure. The *actual* neurons and spike trains in the living organism do not reconstruct environmental stimuli within the organism using these fabricated decoding procedures, or at least the neuroscientist has no evidence of this. If she supposes that other neurons have the function of performing the decoding procedures that she discovered, and she wishes to find biological evidence, then she must record from neurons that allegedly perform the decode, and, using similar mathematical techniques above, fabricate a secondary decoding procedure that links these spike trains to the original decoding procedures. These investigations lead to an infinite experimental regress that mirrors the epistemological regress of the homunculus argument. The only way to stop the regress is to discover the brain-image of the stimulus.

Eliasmith and Anderson (2002) have suggested another way to address this experimental regress when considering mathematically fabricated decoding procedures used in neuroscience, but their solution is quite similar to sweeping the problem under the rug:

> In fact, according to our account, there is no directly observable counterpart to these optimal decoders. Rather, the decoders are 'embedded' in the synaptic weights between neighboring neurons. That is, coupling weights of neighboring neurons indirectly reflect a particular population decoder, but they are not identical to the population decoder, nor can the decoder be unequivocally 'read-off' of the weights. (quotes in original, p. 17)

So long as one can construct a mathematical heuristic to statistically map neuronal activity onto stimuli—which we can always do if the activity is correlated in some way, and becomes more likely if we consider a population of neurons—then we should also assume that the mathematical heuristic is unobservably 'embedded' within synapses. This concept of embeddedness is even more mysterious than representation in that we cannot, even in theory, consistently map synaptic weights to the mathematical decoder. We can avoid the regress, but at the expense of an unfalsifiable and seemingly unscientific assumption.

Application of Shannon's model to neuroscience appears to require embedded decoders and embedded brain-images, both of which are beyond empirical investigation, so the very presence of an encoded message within the brain presents a problem. In other words, why should the brain contain encoded messages that transmit environmental messages, yet never reproduce the structure of the message itself? The organism requires the actual message, and not only an encoded version of it. At this point our analogy to

Shannon's communication model breaks down. It does not appear that the environment communicates a message to the organism, but rather, the organism is perhaps translating the environment. Spike trains are not signals corresponding to encoded messages; they are the actual messages only in the language of the organism, whatever that might mean. With respect to the organism, the message is not encoded in anyway, and speaking of a neural code is metaphorical and at times misleading. The analogy has changed from information transmission to language translation. But even the idea that spike trains are a language is metaphorical—spikes trains need not constitute a private biological language. Our goal here, however, is not to support other metaphors, but to show that Shannon's communication model, which is an integral part of modern technology, does not match the way in which the environment 'communicates' with the organism.

4. The neural code

It is assumed in the neurosciences that there exists something called a neural code, and that this code has something to do with how the brain represents physical properties (and mental properties if you believe in such things). One reason for this informational code talk is the assumed straightforward analogy between the action potentials of neural systems and the classical communication model of source, encoder, channel, decoder, and receiver (Bergstrom and Rosvall 2009), but this analogy is incomplete and generates ambiguities. The decoder and receiver in the model have nothing to do with the organism, but it is the organism that allegedly decodes and receives the message.

The first stage of encoding, in which the properties of an ES are supposedly encoded into patterns of action potential, is probably the least controversial, but on what ground can we call this a code in more than a statistical sense? Compare this code to the genetic code. It is generally agreed that genes code for the amino acid sequences of proteins. The two *alphabets* of the genetic code are finite and well-defined, consisting of the four-letter set of nucleic acid bases (A, G, T, C) and the 23 amino acids plus a stop codon. There is a regular mapping between nucleic acid bases and amino acids, where ordered sequences of bases map to sequences of amino acids. The mapping is one-way in that transcription and translation decode base sequences into proteins, but there is no cellular mechanism that encodes proteins as sequences of bases. We might call this an 'encodeless' code. Still, even authors skeptical of informational talk in biology see reason to call the genetic code a code in a way that means more than causal correlation (Godfrey-Smith 2000; Griffiths 2001).

The two alphabets of the neural code are taken to be environmental physical properties and temporal patterns of neuronal activity. Both alphabets are presumably uncountably infinite because both concern continuous variables, although one can make the case that perception is not infinitely fine-grained and that continuous magnitudes are discretized. While infinite codes exist and neuroscientists have demonstrated how neuronal activity can be used to *estimate* continuous parameters of a stimulus, it remains difficult to clearly specify the alphabets of the neural code (Eliasmith and Anderson 2002), especially in comparison to the genetic code—which strangely has taken more criticism with respect to informational talk than neural coding.

With regard to neural coding and the nature of the code's alphabet, photoreceptors, the first stage or input element of the visual system, bring up several questions. Photoreceptors clearly play an important role in our capacity for vision. Photons from the environment are absorbed by light-sensitive photopigments of the photoreceptor, which through a cascade of biochemical reactions, hyperpolarize the photoreceptor, decreasing the rate of glutamate release from the photoreceptor synapse. The rate of glutamate release is thus graded from its highest rate in complete darkness to lower rates with increasing light absorption and hyperpolarization. Post-synaptic bipolar cells respond to these glutamate levels by generating action potentials.

Should we conclude that the photoreceptor processes code-like information, or is it simply part of a well-specified causal cascade going from photon absorption to glutamate release? Although we can always talk as if the photoreceptor is processing information—in the statistical, correlation sense—we are less drawn to do so, presumably because the mechanistic details of the photoreceptor are fairly-well understood. Either the photoreceptor is processing information, in which case we see that information

processing talk tends to drop out once we sufficiently understand the causal mechanism, or it is not processing information, requiring us to explain why ganglion cells process information but photoreceptors do not. Neuroscientists have tried to cash-out this difference through the dichotomy of implicit versus explicit representations, but implicit here simply means that a successful mathematical decoding heuristic has not yet been devised by the scientist (deCharms and Zador 2000), making the distinction relative.

Next consider glutamate release at the photoreceptor synapse. The rate of glutamate release, when looking at the complete population of photoreceptors, appears to encode everything we need to know about the external visual stimulus. In other words, if we were able to measure the rate of glutamate release at every photoreceptor, then a clever neuroscientist would be able to reconstruct, or decode, the stimulus that led to that particular pattern of glutamate release. The photoreceptor is the encoder, but it encodes physical properties into rates of glutamate release. Those knowledgeable about Shannon information theory will even tell us that the downstream electrical activity resulting from the glutamate release can only degrade the information in the source signal via the so-called data processing inequality. It is therefore reasonable to conclude that the population-rate of glutamate release is an appropriate alphabet of the visual neural code, a code that has nothing to do with action potentials.

In addition to ambiguities in specifying the alphabets of the neural code, there are problems with the 'grammar' of neural coding. An important aspect of all engineered codes is that concatenated sequences or combinations of the 'source' alphabet can be recovered as sequences or combinations of the 'target' alphabet. This combinatorial property is manifest in the genetic code: sequences of bases are systematical mapped to sequences of amino acids. Do neural codes, at least the ones discovered by neuroscientists, unambiguously manifest this property? I will first concede that neuroscientists have demonstrated that *temporal* sequences of action potentials can be recovered as temporal sequences of distinct stimuli. My concern, however, is that neuroscientists have not demonstrated that multiple *contemporaneous* source messages are encoded by any obvious function of the individual message encoding schemes. To be clearer, if S1, S2, and S3 are stimuli and R1, R2, and R3 are spike trains, then given individual encoding relations S1→R1, S2→R2, and S3→R3; does S1+S2+S3→R1+R2+R3, where S1S2S3 occur contemporaneously?

This situation is particular relevant to vision and how neuroscientists understand the receptive fields of cortical cells like those discovered by Hubel and Wiesel. Individual cells, such as complex cells, are classically characterized by their receptive fields and typically fire most when the stimulus is a suitably oriented bar. This leads us to believe there is a clear encoding relation between oriented bars and cellular responses that is always followed, but experiments like those of Bakin et al. (2000) show us that this is simply not the case. In Fig. 3, the dashed-boxes below the x-axis represent the identified receptive field of a striate cortical neuron, and the bar within the dashed-box in the S1 column represents a bar in the preferred orientation for that neuron. S1 is the 'optimal' stimulus for this neuron, and we see for S2 that a bar outside of the receptive field generates little response. However, the simultaneous presentation of S1 and S2 yields a super-optimal response, even though we might expect S2 to not influence the response at all (because it is outside of the classical receptive field). In symbols, S1→R1 and S2→R2, but S3=S1+S2 does not map to any obvious linear function of R1 and R2.

We can appreciate this potential difficulty in neural coding when we consider again the classical stimulus-response curve Fig. 1 as a function R(S), where R is the neural response and S the stimulus. Now consider a response function R(S1, S2) for the same neuron. When each stimulus S1 and S2 is a visual stimulus, R(S1, S2) ≠ R(S1)+R(S2) for some stimuli as above. Since R(S1, S2) is not linearly separable, nor obviously separable by some other transformation, then to understand the encoding relation R(S1, S2), we must experimentally determine this relation by probing the neuron with a wide range of joint stimuli S1+S2. A similar procedure is required to determine R(S1, S2, S3) and so forth, procedures that quickly become impractical for moderate numbers of stimuli, yet if the piecewise encoding relations do not clearly determine combinatorial relations, then in what sense is the piecewise code a code in the first place? Scientists may

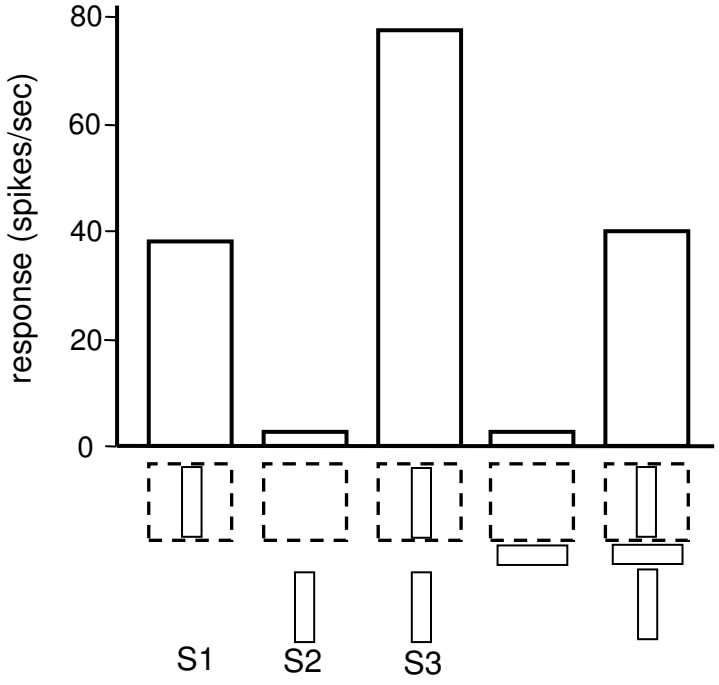

Fig 3. Response of a neuron in the primary visual cortex. Dashed boxes represent the receptive field of the neuron. A bar in the middle of the dashed-box is the optimal stimulus for the neuron, while a bar outside of the box is a stimulus that does not activate the neuron. The combination of a bar inside and outside of the receptive field produces a super-optimal response, which is not predicted by either stimulus taken in isolation. Adapted from Bakin et al. (2000)

use nonlinear mathematical techniques to patch together piecewise encoding relations into combinatorial relations, but then, as in the previous section, one may question the biological relevance of such a practice.

As well as lacking contemporaneous combinatorial properties, responses of neurons in primary and extrastriate cortex are not strictly determined by the passive physical properties of external stimuli, but are dependent upon the behavioral and cognitive states of the organism (Pasternak et al. 2003). These sorts of dependencies are not present in the genetic code—at least not unless one stretches code-like talk beyond the coding of proteins and onto whole-organism phenotypes, which does not appear appropriate (Godfrey-Smith 2000; Griffiths 2001). One might argue that behavioral dependences are 'part of the code', making the encoding relation look something like (S, B)→R where B is a behavioral or cognitive state, but this move greatly complicates any understanding of the alphabet and grammar of the neural code. What behaviors or cognitive states count in the code? How do we specify the relevant alphabet? Do (S, B) pairs have combinatorial properties, are not behaviors functions of neural responses, and will mappings not change with time and learning? Again, one can respond that "it's complicated, but this is what takes to understand the neural code", yet this code looks less and less like a systematic code and more like an unruly temporally dependent multidimensional blob of lawless correlations.

Neural coding, as practiced by cognitive scientists and neuroscientists, is an anything-goes sort of coding, by which I mean that any brain properties that correlate with stimulus properties, behavioral properties, cognitive properties, or social contexts can be considered part of the code. This is not a shortcoming or limitation of neuroscientific research, but it is a limitation of the coding metaphor. The code of the neural code does not have a lot going for it when critically compared to genetic or engineered codes. One property they probably share is that their respective mapping relations are arbitrary. Garson (2003) has attempted to defend a concept of information as applied to neural responses based upon this property. Given the above shortcomings of neural coding, it is difficult to see how arbitrariness by itself justifies informational and code-like language.

5. Information theory as a tool

The decoding procedures discovered by neuroscientists are useful in that they allow us to predict spike trains given environmental stimuli, and stimuli given spike trains; but the specific decoding procedures do not tell us anything about the function of neuronal populations—because the decoding algorithms have nothing to do with the biology of the organism. Rather, the capacity to successfully predict between stimuli and spike trains via decoding is typically taken as evidence that spike trains represent stimuli, although the capacity to predict immediately

follows from the statistical correlations between spike trains and stimuli.

There are neuroscientists who consistently, and with clearly stated assumptions, apply Shannon's mathematical information theory to neuronal data with the goal of quantifying the theoretical channel *capacity*, or bit rate, of spike trains (Strong et al. 1998; Reike et al. 1999). These interesting applications of information theory within neuroscience try to answer the following question: assuming spike trains carry Shannon information about the environment, how much information (in bits) could they carry? We could ask similar questions about the oxygen molecules in one's living room, the ants in an anthill, or the blades of grass in one's yard—although the answers presumably would not be as interesting. The fact that Shannon information theory can be rigorously applied to spike trains does not imply that the brain processes information as a function.

Other neuroscientists, such as deCharms and Zador (2000), repeatedly claim that spike trains carry information about the environment as a fact, and suggest what it means to carry information: "Imagine recording from the neuron labeled B1 during different types of stimuli or behaviors and discovering the information that this neuron carries about the organism's environment—the content of this neuron's signal" (p. 614-15). In a concrete example about a retinal cell they say that "The activity of the neuron will be highly correlated with the point of luminance (thus carrying content about this input)" (p. 637). Like in Hubel-Wiesel's ES/BR experiments, we call this evidence the selective covariation between stimulus properties and spike trains. deCharms and Zador use the word 'information' above to possibly mean 'specific properties or features of the stimulus.' Given these examples, we can suppose that they would endorse the following argument: (1) spikes trains and stimulus properties selectively (and causally) covary, and (2) the (representational) content of a spike train is the stimulus property that causes that spike train.

deCharms and Zador do not bring forth any other types of experimental evidence other than selective covariation to justify the claim that spike trains carry informational or representational content, although they do stress that the representational nature of spikes trains is based upon content and function. We have argued that (1) is a statement about the evidence that all of us would agree upon, but that (2) does not obviously follow. The fact that an ES and BR selectively covary, through causal paths, does not appear sufficient to justify claims of representational content, and it has been argued that covariation of this sort is not even necessary for representational content (Millikan 1989; Bechtel 1998).

We need not expect deCharms and Zador, as neuroscientists, to philosophically justify what it means for a spike train to carry representational content, yet if claims of carrying content do not follow immediately from the observed evidence, then we can only assume that they are interpreting the evidence or communicating the evidence by way of metaphor. But deCharms and Zador, along with many other neuroscientists, speak as though 'carrying content' is a straightforward experimental fact apart from, or in addition to, selective correlations.

To justify informational talk, some neuroscientists mount a proof-is-in-the-pudding defense, arguing that the use of informational concepts has helped the field of neuroscience progress. How else could visual neuroscience be where it is now without envisioning hierarchies of different layers of neurons that process specific types of information? I believe confusion arises between representations of experimental facts—which we use to summarize and share our findings with others—and taking experimental facts to be representations. For example, there is a tendency to imagine that a receptive field (or stimulus-response tuning curve, e.g. Fig. 1) is a property of a neuron, but this is no straightforward *intrinsic* physical property of a cell. If anything a receptive field is a mathematical function of physical properties spread throughout the brain and includes, from the start, reference to external stimuli. When a receptive field is first characterized, it simply represents the collection of stimulus-response pairs (in the form of a handy graph or picture) that were investigated by the researcher. The receptive field summarizes the results of an experiment involving a particular neuron; it is not an intrinsic physical or biological property of that neuron.

These representations of experimental facts are undoubtedly useful in guiding further research and

generating hypotheses. Neuroscientists can and do reason using these representations, but this does not imply that, for instance, spike trains carry representational content. Similarly, Shannon information theory can be usefully applied to neuronal responses, but that does not imply that neuronal responses are code-like or process information, at least no more so than any objects that manifest causal correlations. We appear to be confusing the tools we use to understand neural systems with the properties of neural systems. An appreciation of this distinction may help us to better understand the function of neurons and the brain.

References

Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings. Part II: the response of a single end organ. *The Journal of Physiology*, 61, 151-171.

Adrian E. D., & Matthews, R. (1927). The action of light on the eye. Part I. *The Journal of Physiology*, 63, 378-414.

Amundson, R., & Lauder, G. V. (1994). Function without purpose: the uses of causal role function in evolutionary biology. *Biology and Philosophy*, 9, 443-469.

Bakin, J. S. , Nakayama, K., & Gibert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *The Journal of Neuroscience*, 20, 8188-8198.

Bechtel, W. (2006). *Discovering cell mechanisms: the creation of modern cell biology.* Cambridge: Cambridge University Press.

Bechtel, W. (2008a). *Mental mechanisms: philosophical perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W. (2008b). Mechanisms in cognitive psychology: what are the operations? *Philosophy of Science*, 75, 995-1007.

Bechtel, W., & Richardson, R. C. (2010). Neuroimaging as a tool for functionally decomposing cognitive processes. In S. J. Hanson, & M. Bunzl (Eds.), *Foundational issues in human brain mapping* (pp. 241-262). Cambridge: MIT Press.

Bergstrom, C., & Rosvall, M. (2009). The transmission sense of information. *Biology and Philosophy*, 26, 159-176.

Bigelow, J., & Pargetter, R. (1987). Functions. *The Journal of Philosophy*, 84, 181-196.

Bunzl, M., Hanson, S. J., & Poldrack, R. A. (2010). An exchange about localism. In S. J. Hanson, & M. Bunzl (Eds.), *Foundational issues in human brain mapping* (pp. 49-54). Cambridge: MIT Press.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). New York: Wiley.

Cummins, R. (1975). Function analysis. *The Journal of Philosophy*, 72, 741-765.

deCharms, R. C., & Zador, A. (2000). Neural representations and the cortical code. *Annual Review of Neuroscience*, 23, 613-647.

Dretske, F. (1981). *Knowledge and the flow of Information.* Cambridge: MIT Press.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge: MIT Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural Engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge: MIT Press.

Garson, J. (2003). The introduction of information into neurobiology. *Philosophy of Science*, 70(5), 926-936.

Garson, J. (2011). Selected functions and causal role functions in the brain: the case for an etiological approach to neuroscience. *Biology and Philosophy*, doi 10.1007/s10539-011-9262-6

Godfrey-Smith, P. (2000). On the theoretical role of 'Genetic Coding,' *Philosophy of Science*, 67, 26-44.

Godfrey-Smith, P, & Sterelny, K. (2008). Biological Information. In E. D. Zalta (Ed.) *The Stanford*

Encyclopedia of Philosophy (Fall 2008 ed.). http://plato.stanford.edu/archives/fall2008/entries/information-biological/. (Accessed 7 April 2011)

Griffiths, P. E. (2001). Genetic Information: a metaphor in search of a theory. *Philosophy of Science*, 68, 394-412.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge: University of Cambridge Press.

Hardcastle, V. G., & Stewart, C. M. (2002). What do brain data really show? *Philosophy of Science*, 69(3), S72-S82.

Henson, R. (2006). Forward inference using functional neuroimaging: dissociations versus associations. *Trends in Cogntive Sciences*, 10(2), 64-69.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106-154.

Kitcher, P. (1993). Function and design. *Midwest Studies in Philosophy*, 18, 379-397.

Klein, C. (2010). Images are not the evidence in neuroimaging. *British Journal for the Philosophy of Science*, 61(2), 265-278.

Logothetis, N. K., & Brain, A. W. (2004). Interpreting the BOLD signal. Annual Review of Physiology, 66, 735-769.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453, 869-878.

Maynard Smith, J. (2000). The concept of information in biology. *Philosophy of Science*, 67, 177-194.

Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge: MIT Press.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, 86(6), 281-297.

Neander, K. (1991). Functions as selected effects: the conceptual analyst's defense. *Philosophy of Science*, 58, 168-184.

Pasternak, T., Bisley, J. W., & Calkins, D. (2003). Visual information processing in the primate brain. In M. Gallagher, & R. J. Nelson (Eds.), *Biological Psychology* (pp. 139-185). New York: John Wiley & Sons Inc.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics*, 37, 1-38.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59-63

Popper, K. R. (1959). The *Logic of Scientific Discovery* (translation of *Logik der Forschung*). London: Hutchinson.

Rieke, F., Warland, D., van Steveninck, R., & Bialek, W. (1999). *Spikes: exploring the neural code*. Cambridge: MIT Press.

Roskies, A. L. (2007). Are neuroimages like photographs of the brain? *Philosophy of Science*, 74, 860-872.

Rudwick, M. J. S. (1964). The inference of function from structure in fossils. *British Journal for the Philosophy of Science*, 15, 27-40.

Sarkar, S. (1996). Decoding "coding" — information and DNA. *BioScience* ,46, 857-864.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.

Strong, S. P., Koberle, R., van Steveninck, R., & Bialek, W. (1998). Entropy and information in neural spike trains. *Physical Review Letters*, 80, 197– 200.

Uttal, W. R. (2001). *The new phrenology: the limits of localizing cognitive processes in the brain*. Cambridge: MIT Press.

Wright, L. (1973). Functions. *Philosophical Review*, 82, 139-168.