Neural Information and the Problem of Objectivity

Charles A. Rathkopf

To appear in *Biology and Philosophy*

Abstract

A fascinating research program in neurophysiology attempts to quantify the amount of information transmitted by single neurons. The claims that emerge from this research raise new philosophical questions about the nature of information. What kind of information is being quantified? Do the resulting quantities describe empirical magnitudes like those found elsewhere in the natural sciences? In this article, it is argued that neural information quantities have a relativisitic character that makes them distinct from the kinds of information typically discussed in the philosophical literature. It is also argued that despite this relativistic character, there are cases in which neural information quantities can be viewed as robustly objective empirical properties.

Consider the claim that the H1 neuron in the visual system of the blowfly transmits information at a rate of 81 bits per second. This claim conveys a typical result generated by a fascinating research program in neurophysiology that attempts to estimate the rate at which information flows through individual neurons (Rieke et al., 1999; Frye and Dickinson, 2001; Van Hateren et al., 2005; Neri, 2006). Working out exactly what is meant by such claims leads rapidly, and inexorably, into philosophically contentious terrain. Despite this, such claims have received hardly any serious attention in the philosophical literature.

Here I attempt to answer two questions about quantitative estimates of neural information. First, what kind of information do such estimates purport to describe? Taxonomies of informational concepts have been developed both in the philosophy of mind and in the philosophy of biology. It is unclear, however, that the kind of information referenced in the claim above conforms to any of the existing analyses. The second question is whether information rate claims describe objective empirical magnitudes like those we find elsewhere in the natural sciences. Analogical reasoning provides prima facie grounds for doubt. If I send you a message in Morse code, and you happen not to know Morse code, there is a sense in which little or no information has been transmitted, regardless of what the decoded message might have said. In that sense, the *quantity* of information transmitted by a physical signal depends on the interpretive capacities of a receiver, and consequently appears to lack a certain kind of objectivity.

Of course, the sort of information described in the claim about the blowfly H1 neuron is quite unlike the kind of information typically transmitted by means of conventional human symbol systems. Nevertheless, the comparison is not entirely void of interest. In what follows, I'll argue that there is a subtle but theoretically significant sense in which the quantity of information transmitted by a neural signal is relative to the capacity of a receiver mechanism to make use of the signal. In order to answer the two questions above, it is necessary to understand that relativity. I'll begin by discussing a simple model to illustrate what receiver-relativity means in a quantitative setting. Then, I'll argue that receiver-relativity is a real feature of the empirically driven estimates of neural information transmission. In the second half of the article, I'll build upon the discussion of receiver-relativity to develop answers to the two questions posed in the previous paragraph.

1 Receiver-relativity in a simple model

Imagine a nocturnal organism with an extremely simple sensorimotor arc that drives locomotion. There are three variables to consider. First, there is an environmental property X, which represents luminance. As far as the organism is concerned, X can take on only three states, *bright*, *dusk*, and *dark*. Second, there is a single perceptual neuron, Y, which can take on three discrete states α and β, and γ. Third, there is a motor neuron, R, which controls locomotion. The states of R are driven by the states of Y. The coupled system XYR is a communication device in the sense associated with the mathematical theory of communication. (Shannon and Weaver, 1949). As Figure 1 shows, X plays the role of the information source, Y plays the role of the information transmitter, and R plays the role of receiver. The fourth element in the diagram, which Shannon and Weaver called the destination, is here interpreted somewhat abstractly as the behavior of the organism that results from the motor signal at R.

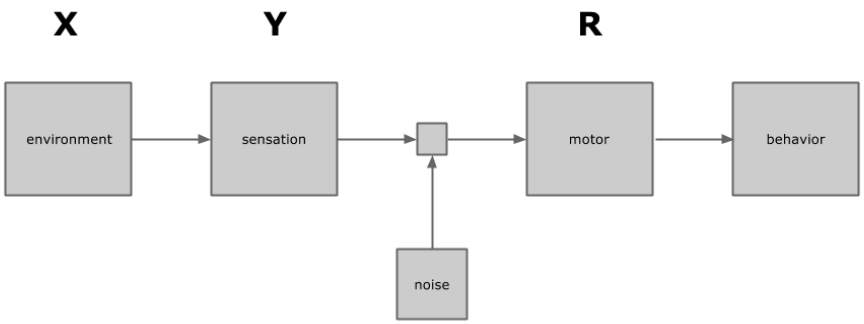

Figure 1: This figure is a neurophysiologically-oriented interpretation of the classic diagram of a communication system that appears in Shannon and Weaver (1949). The labels X, Y, and R correspond to what they called the source, the transmitter, and the receiver, respectively. The rightmost box corresponds to what they called the destination.

This is not the only way to interpret the neurobiology of perception in terms

of Shannon's diagram, but this interpretation provides a setup in which we can ask questions about how the quantitative relation between two variables might be influenced by a user of that relation. This way of setting up the XYR model is reminiscent of an analogy that is common in discussions of neural representation: that of reading a geographic map. X corresponds to the terrain, Y corresponds to the map, and R corresponds to the person who uses the map in order to navigate the terrain. The markings on the map are correlated with the features of the terrain. It is in virtue of her ability to exploit that correlation that the map-reader can navigate successfully. Similarly, in the XYR model, the organism is able to decide whether to move or stay still by attending to the correlation between X and Y.

	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$	X_m
bright	.5	0	0	.5
dusk	0	.25	0	.25
dark	0	0	.25	.25
Y_m	.5	.25	.25	

Table 1: A relative frequency table which represents a hypothetical joint probability distribution over X and Y. The rightmost column represents the marginal distribution of X, while the bottom row represents the marginal distribution of Y. Where I is the mutual information and H is the entropy, $I(X;Y) = H(X) + H(Y) - H(X,Y)$. The upper bound on the amount of transmittable information is given by the lesser of the two marginal entropies. Here, the mutual information is: I(X;Y) = H(X) + H(Y) - H(X,Y) = 1.5 + 1.5 - 1.5 = 1.5 bits/message.

Table 1 describes a hypothetical joint distribution over X and Y. Following Shannon and Weaver, I refer to the states of X as *messages*, and to the states of Y as *signals* (Shannon and Weaver, 1949, p. 2). The set of signals and the set of messages are each represented by a random variable that follows a given probability distribution. (An alternative but mathematically equivalent formulation: the messages are the symbols *sent from* X to Y; the signals are the symbols *sent from* Y to R.) Assuming that all messages and signals are elementary symbols, (i.e. we ignore combinatorial codes) a simple calculation shows that the mutual information between X and Y is 1.5 bits per message.[1]

[1]The entropy associated with a single value of some random variable is given by the log of the reciprocal of its relative frequency. $\text{Log}_2(1/.5) = 1$ bit. $\text{Log}_2(1/.25) = 2$ bits. To find the entropy of an entire distribution, we take a weighted sum over all individual entropies.

Note that nothing thus far has been said about R. Can the properties of R influence the amount of mutual information instantiated by the XY relation? To explore this question, let us imagine that R is completely and chronically insensitive to the distinction between β and γ. What consequences flow from this supposition?

There are two principled ways of answering this question. According to the first answer, the XY relation is to be viewed as a simple physical relation that is not causally influenced by R, which, after all, is located downstream in the causal chain. On this view, we will have to say that that although 1.5 bits of information are indeed transmitted from X to Y, that 1.5 bit quantity is not directly relevant to the functional capacities of the organism. Because the receiver mechanism can discriminate only a portion of the underlying distribution, only a portion of the underlying distribution is relevant to explaining how the organism manages to achieve behavioral control. The rest is explanatorily idle. On this view, quantities of neural information can be assessed in isolation from questions of biological function.

According to the second answer to our question about how R might influence the XY relation, there is a flaw in the way we have attempted to describe the scenario thus far. What sort of flaw? Notice that because the mutual information between X and Y is logically entailed by the underlying probabilities, the only way that R can influence the mutual information is by influencing those probabilities. According to this second way of thinking, R's insensitivity to β and γ demonstrates that the given distribution is not an accurate representation of the biological facts. If the organism cannot, even in principle, exploit the XY relation for some biological end, then the correlation expressed by that relation isn't one that can legitimately be used to compute the mutual information between X and Y.

This second way of looking at the scenario is motivated by the thought that neural information is *essentially* an expression of the biological capacities of the organism. According to this second view, if we want to compute I(X,Y) accurately, we must ensure that the given distribution include only those values that have biological relevance. How can this restriction be accommodated in

So, the marginal entropy H(X) = .5(1) + .25(2) +.25(2) = 1.5. The computation required to find the marginal entropy H(Y) is identical to that required for H(X). To compute the joint entropy H(X,Y), we take a weighted sum over the individual entropies associated with each of the six terms in the center of the table. Three of those terms evaluate to 0. Once they are removed, the remaining terms constitute an expression that is identical to that for H(X), which, as we just saw, is equal to 1.5 bits.

our procedure for estimating I(X,Y)? The accommodation is straightforward enough: since R is insensitive to the distinction between β and γ, we rewrite the relative frequency table so that β and γ are counted together as two instances of one biologically exploitable variable. In that case, a simple calculation shows that there is exactly 1 bit of information flowing through the XYR system.

If this second way of looking at the scenario is correct, the original quantity of 1.5 bits is a kind of fluke. A couple of paragraphs back, I said that the 1.5 bit quantity was explanatorily idle. In light of that description, we might dub the correlation from which it was computed an *idle correlation*. This notion can be contrasted with the more familiar notion of a spurious correlation. A spurious correlation between factors A and B is one in which there is no direct causal relationship (in either direction) between A and B. Either the AB relation is accidental, or it is the result of a common cause. An idle correlation need not be spurious. The XY relation *is* a direct causal relation, and the correlation evident in Table 1 is perfectly real. The idleness of an idle correlation stems rather from the fact that the neural pathway in which it is embedded doesn't seem to "care about" it. For that reason, it fails to constitute an empirically adequate explanatory factor.

We can extract two lessons from the comparison between our two interpretations of the insensitivity scenario. The first lesson is about how to explicate the concept of receiver-relativity. An information system is receiver-relative just in case the mutual information between the set of messages and the set of signals depends on facts about how the underlying correlation is exploited by downstream mechanisms. The second lesson concerns the evaluation of our two ways of looking at the insensitivity scenario. The choice between them seems to turn on the issue of biological relevance. If neural information quantities express purely physical facts that float free from considerations of biological function, then the first way of looking at the scenario makes good sense. If, however, neural information quantities express facts about the functional capacities of an organism, then the receiver-relative view is more appropriate.

2 Some strengths of the receiver-relative view

In this section I argue that the receiver-relative conception of information does a better job of capturing the content of information rate claims in neurophysiology than does the non-relative conception.

Someone might argue for the opposite view as follows. No matter how the organism behaves, as long as there is a way of describing the X and Y variables such that the probability distribution in Table 1 accurately represents their activity, the 1.5 bit quantity follows with necessity. The best response to this claim is not to deny its truth, but its relevance to scientific theory. There is indeed a coherent notion of information according to which the XY relation expresses 1.5 bits, but it is not the notion of information we should be interested in if we hope to learn anything significant about biology. Insisting on the correctness of the 1.5 bit quantity is like insisting that there is a lot of information latent in the correlation between the hair on your head and the direction of the wind. Since no mechanisms are designed to make use of that correlation, the sense in which it carries information is not the sense we have in mind when we say that a neuron has transmitted a particular number of bits. When we say that a neuron has transmitted a particular number of bits, what we mean is that it has transmitted a particular number of *useable* bits. [2]

To support this claim, I want to highlight the role of the signal-noise distinction in experimental practice. The first thing to note is that probability distributions over stimulus and response are never simply given to us, as was presumed in the XYR model. The experimental strategy is not to first examine the probabilities, and then, in a second step, determine what portion of the signal distribution has biological relevance. Rather, experiments are designed to capture biologically relevant signals directly, so that the observed data are already sorted into signal and noise components.

To see how this works, consider the following expression for mutual information. Although it is mathematically equivalent to the one used above, this new expression has the benefit of more closely mirroring experimental operations.

$$I(X,Y) = H(Y) - H(Y|X) \tag{1}$$

The H(Y) term on the right side refers to the full entropy of the perceptual neuron, and it is operationalized by presenting the organism with a wide variety of random stimuli, which, at least in theory, will elicit a representative sample of the full range of physiologically possible response rates. H(Y|X) is

[2] It is worth noting here that the hair-in-the-wind argument exploits a perfectly contingent fact about humans. Filiform hairs on the legs of crickets move with the local air currents too. But in that case, neural receiver mechanisms use the hair-air correlation for predator detection (Magal et al., 2006). In that case, hair direction really is an informational signal. At least in principle, the amount of information transmitted in this case could be quantified.

called the *neural noise* (Dayan and Abbott, 2001, p. 74; Borst and Theunissen, 1999, p. 950.) It is typically measured by presenting the organism with repeated instances of the neuron's preferred stimulus. The idea behind this operationalization is that, when a given value of X is simply repeated, there isn't any variation in its value about which one could hope to be informed. Under these conditions, Y is not transmitting any information, despite the fact that it remains active. Crucially, the set of responses elicited under these non-functional conditions is not entirely abnormal. This spontaneous activity will reappear under conditions in which the stimulus at X really does vary. Since we know this activity is non-functional, its contribution to the variation must be filtered out.

The idea that the H(Y|X) term should be regarded as non-functional activity is equivalent to the assumption that the function of Y is to report to downstream mechanisms on the status of X-like stimuli. This assumption shapes the experimental design needed to get an accurate measurement of the information rate. If, in fact, the function of Y is something other than reporting on the nature of X-like stimuli, then we would need to employ a different experimental setup, with different stimuli and background conditions. So we must ask: what empirical facts make the actual measurement setup the right one to employ? The actual setup is the right one because, as a matter of empirical fact, Y-signals are *used* by the flight motor for the purpose of navigating through stimuli of precisely the kind employed in the experiment.

This is the crucial insight behind the argument that informational quantities are receiver-relative. The nature of Y's function depends on the manner in which its activities are exploited by downstream mechanisms. Since experimental procedures show that the value of I(X,Y) depends on Y's functional role, we can say that the value of I(X,Y) depends on how signals from Y are used downstream. And since, according to the definition in Section 2, an informational quantity is receiver-relative just in case the quantity depends on how the underlying correlation is used, I(X,Y) is a receiver-relative quantity.

To make this more concrete, notice that the assumption that the unique function of Y is to report to downstream mechanisms on the nature of X-like stimuli is far from trivial. Many neurons play functional roles that are far too subtle to be isolated in behavioral experiments. This is one reason why experiments are only performed on a small number of simple model systems and why those model systems are almost invariably perceptual or sensory. Only systems with a highly streamlined functional profile which is systematically related to

environmental properties are sufficiently transparent for experimentalists to confirm that a particular experimental setup is appropriate for getting an accurate estimate of I(X,Y).

Take the H1 neuron as an example. It is probably fair to say that the H1 neuron is the chief model neuron in information-theoretic neurophysiology. Why is that? H1 makes for a great model neuron because the simplicity of the system in which it is embedded makes it possible to reliably detect when its signals have been received successfully. When H1 signals are received successfully, they influence the flight of the organism quite directly. Even in this simple system however, confirming that a particular experimental setup accurately exploits the biological function of the neuron requires a rich patchwork of background knowledge. First, lesion experiments have established that H1 signals track horizontal optic flow (Frye and Dickinson, 2001). On the basis of that knowledge, experimentalists can be confident that vertical sine-wave gratings will serve as appropriate stimuli, and that fine tuning the contrast and angular velocity of such stimuli will trigger something close to the maximal neural response. Second, the perception-action loop from H1 to the flight motor is tight. Each H1 cell synapses with centrifugal horizontal cells that govern the part of the flight motor that creates horizontal torque (Neri, 2006). This is important because it means that the successful receipt of H1 signals can be confirmed through behavioral observation. Moreover, there is only one H1 neuron in each lobule of the fly's brain; one corresponding to each eye. This is strong anatomical evidence that the H1 is the *only* perceptual neuron that contributes in such a direct way to horizontal flight control. As a result, whenever we observe the fly making highly accurate horizontal flight adjustments (which it does at a millisecond resolution, flying at speeds up to two meters per second), we can be sure that the signals responsible for that behavior are coming from H1 (Frye and Dickinson, 2001). Without this rich patchwork of behavioral and anatomical knowledge that tells us what H1 signals are supposed to accomplish, we would not be able to tell whether the activity of the H1 neuron is being exploited on any particular occasion. And if we didn't know that, we wouldn't know whether our experimental design accurately captures the noise in the system. The take home point here is that receiver-relativity is not only real, it is consequential. We cannot reliably estimate neural information quantities without (i) having a clear understanding of how the underlying stimulus-response correlation is exploited to achieve some biological end, and (ii) confirmation that the experimental setup

actually exploits that function.[3]

3 What kind of information is being estimated?

We are now in position to investigate the first of the two questions posed in the introduction: what kind of information do estimates of informational quantities in the brain purport to describe? We've already established that informational quantities are receiver-relative. Now the goal is provide an analysis of the empirical magnitude that we have quantified, and see how that empirical magnitude fits into the larger landscape of philosophical thought about information.

The question is motivated in part by a desire to better understand the content of neurophysiological theory. But it is also motivated by a desire to better understand informational quantities generally. Informational quantities are philosophically interesting in part because they seem to straddle two radically disparate conceptual arenas. From one perspective, information quantities are the unsurprising result of a choice to represent ordinary empirical phenomena with a particular collection of mathematical tools. Rather than representing an empirical variable with a probability distribution, we represent it as a logarithmic function of a probability distribution. This change in mathematical notation, it seems natural to think, has no deep metaphysical implications at all. The empirical variables that we have elected to represent with this logarithmic notation are no less ordinary than any other feature of the empirical world that lends itself to probabilistic representation. From another perspective, however, the appropriateness of information-theoretic tools reflects the fact that there is something signal-like about the character of the empirical phenomenon itself. We employ information-theoretic tools precisely because we hope to highlight that signal-like character. It is from this perspective that it seems fitting to characterize information as the currency of communication. From this perspective, there is indeed something unusual about the empirical magnitudes we attempt to quantify information-theoretically. To see this, one must only recall the truism that communication is, at least paradigmatically, a relation between

[3]I have not given a definition of the term "function." As the size of the philosophical literature on the subject suggests, it is not easy to say exactly what it means for an object or process to have a biological function. For many aspects of biological theory, including the kind of function discussed here, I favor the modern history theory of functions, as expressed in Godfrey-Smith (1994). But my view is compatible with other theories of biological function as well. It is important, however, that the notion of function have some relation to natural history. Without that connection, it loses some of the objectivity that I argue is worth retaining in neurophysiological theory. See Section 4 for further commentary on this point.

agents. With that truism in mind, the treatment of small and patently non-thinking neurons in informational terms is deeply interesting, and maybe even perplexing.

Both in the philosophy of mind and in the philosophy of biology, this tension between the ordinary and the perplexing aspects of information has been diagnosed as the result of verbal ambiguity. Our word "information," it is often claimed, denotes two distinct concepts. One of them, which, for the purposes of disambiguation is often labeled *Shannon information,* is a technical concept. It may evoke conceptual puzzles related to the interpretation of probability, but its use does not require any controversial metaphysical assumptions about the nature of communication. The second concept, which is often labeled *semantic information*, is taken to be central to understanding communicative phenomena, and is also taken to raise a special set of problems in the philosophy of mind and the philosophy of biology. Many sub-varieties of information have been discussed in the philosophical literature, but every philosophical survey of information introduces this distinction.[4] It is therefore appropriate to restrict the discussion to the question of whether the sort of information under scrutiny in this discussion can be subsumed under either of these two umbrella categories.

3.1 The semantic interpretation

Are neural information estimates best interpreted as claims about semantic information? It is common, especially for sensory neurons, to be described as being engaged in acts of representation. For example, in a passage about the filiform hairs on cricket cerci, Purves et al say "peripheral sensory neurons associated with the hairs represent the full range of air current directions and velocities impinging on the animal" (Purves et al., 2001, p. 195). Since representational phenomena and semantic phenomena are closely related, the semantic interpretation seems plausible.

To investigate the semantic interpretation more carefully, we must first notice that the question we have posed is a bit ambiguous, and can be fruitfully separated into two lines of inquiry. The first is whether the phrase "81 bits/s" in the claim "The H1 neuron transmits information at a rate of 81 bits/s" picks out a semantic property of a neural signal. The answer to this question is a decisive 'no.' The question of how many bits are associated with an informational signal

[4] A very recent survey is Floridi (2015). Other prominent surveys include Sayre (1976), Adams (2003), and Harms (2006).

is orthogonal to the question of what semantic features that signal expresses. Two strings can have entirely different meanings and nevertheless transmit the same number of bits. Consider, for example, a long binary string. The ones and zeros may be equally probable, but that hardly entails that they mean the same thing.[5] Conversely, two tokens of one string may be semantically equivalent but nevertheless transmit different quantities of information. For example, the meaningful string "it is currently raining" carries more information for someone in a windowless room than it does for someone standing outside.

The second line of inquiry prompted by the semantic interpretation of neural information estimates is the following. Do neural information estimates describe how much semantic information is flowing through a circuit, without purporting to specify what the semantic content is? In other words, do neural information estimates measure *how much meaning* is flowing through a system? Mundane examples suggest that this idea is at least coherent. One might want to insist, for example, that the quantity of semantic information conveyed by a particular sentence is less than the quantity conveyed by the book in which it appears. There is an interesting history in logic, tracing back to the work of Bar-Hillel and Carnap (1953), which tries to make this idea precise. One reason to resist this interpretation of neural information estimates is that the amount of semantic information in a system is a matter of what is said about things outside the system, whereas the amount of information in neural information rates is determined entirely by the probabilities attached to the variables that constitute the system. Semantic properties are relational. They hold between a symbol and the thing it stands for. Probabilities need not be relational. They can just describe the frequency with which one kind of thing happens, without talking about the relation between that thing and something else.[6]

[5] I have suppressed the role of time in this discussion because it complicates the mathematics without changing the conceptual issues under consideration. Notice that the argument doesn't change significantly if we consider two non-identical strings sent from one location to another over time. If strings share statistical properties, their transmission may achieve the same *information rate* expressed in bits/s. This is still no reason to think that the two strings have the same meaning.

[6] Another, more controversial, reason to resist the semantic interpretation of information theoretic estimates is that the activity in a single neuron seems to be too low-level for semantic properties to emerge at all. If there are no semantic properties at the level of individual neurons, then, clearly, information estimates describing the behavior of individual neurons cannot be interpreted semantically. Rosa Cao has defended this anti-semantic position on the basis of an interesting dilemma. The signals transmitted by individual neurons either lack sufficiently robust connections to the external world to carry content on their own, or the connections they exhibit are too inflexible too deserve an informational, as opposed to merely causal, mode of description (Cao, 2012).

3.2 The Shannon interpretation

If neural information estimates are not best interpreted in terms of semantic information, what about Shannon information? Here the situation is a bit more difficult to assess because philosophical commentary on the nature of Shannon information has been somewhat out of step with its use in science and engineering. Definitions of Shannon information in the philosophical literature tend to have an ontological orientation that is completely absent in technical manuals on the subject, where definitions of key concepts are purely mathematical, and no suggestions are made about the range of empirical phenomena to which the mathematics is legitimately applied.[7] The ontologically oriented definitions found in the philosophical literature tend to be radically permissive. Consider, for example, the definition proposed in a review paper on information in biology by Godfrey-Smith and Sternly: "For Shannon, anything is a source of information if it has a number of alternative states that might be realized on a particular occasion. Any other variable contains some information about that source, or carries information about it, if its state is correlated with the state of the source"(Godfrey-Smith and Sterelny, 2008).[8] A similar definition is found in a more recent article on information by Piccinini and Scarantino that gives a thorough overview of informational concepts. They say: "The identity of a communication theoretic message is fully described by two features: it is a physical structure distinguishable from a set of alternative physical structures, and it belongs to an exhaustive set of mutually exclusive physical structures selectable with well-defined probabilities" (Piccinini and Scarantino, 2011, p. 20).

As highlighted by their use of the term "identity," the two criteria mentioned in the Piccinini and Scarantino definition are to be read not only as necessary conditions, but also as sufficient conditions for the presence of quantitative information. Godfrey-Smith and Sternly's use of the phrase "any other variable" likewise suggests that correlation between empirical variables is to be read as a sufficient condition. According to such permissive definitions, then, anything that can be modeled by information theory is, ipso facto, information.

Working scientists can be forgiven for asking what the point of such a permissive concept could possibly be. The motivation behind the permissive concept is

[7] See, for example, Cover and Thomas (1991).

[8] The phrase "For Shannon" in this definition might be misleading. Shannon was many things, but he was not a metaphysician. He was not interested in trying to divide the world into informational phenomena and non-informational phenomena. In fact, in a short paper entitled "The Bandwagon," Shannon warns that information theory is easily misused when applied outside the realm of communications technology (Shannon, 1956).

its metaphysical innocence. The desire for metaphysical innocence traces back to the work of Fred Dretske who, in his 1981 book *Knowledge and the Flow of Information*, hoped to develop a naturalistic and reductive theory of semantics (Dretske, 1981). In order for the desired reduction to count as a theoretically significant achievement, the reductive base had to be something far less metaphysically controversial than the phenomenon to be reduced. Since covariation between empirical variables is about as metaphysically innocent as it gets, it was a natural choice for Dretske.

With this background on the philosophical appropriation of Shannon's ideas in place, we can return to the question of whether the philosopher's permissive notion of Shannon information is the right concept with which to interpret neural information estimates. Again, I think the answer is 'no.' Definitions of information focused on covariation between empirical variables are so easily satisfied that they rule out almost nothing. In particular, they cannot distinguish the correlations that support genuine information transmission from merely *idle correlations*. Recall from Section 2 that an idle correlation is one which may be supported by a direct causal relationship, but which is, nevertheless, not among the correlations that are exploited for purposes of biological control.

If we employ the permissive conception of information to interpret the claim that the H1 neuron transmits 81 bits/s, the empirical content of the resulting claim is implausibly sparse. It says, in effect, that we happen to have observed a correlation which led us to the 81 bits/s figure. In fact there is nothing happenstance about it. Experiments were painstakingly designed to evoke the performance capacity of the neuron. As discussed in Section 3, in order to get an accurate estimate of the mutual information, experiments must be designed to filter out idle correlations (noise), from genuine information-supporting correlations (signals). We can conclude that the permissive concept is not the right one for interpreting neural information claims. The concept of neural information, therefore, cannot be subsumed under either of the two prominent umbrella categories employed in the existing philosophical taxonomies of informational concepts.

If the quantitative kind of information described in neurophysiological estimates is neither semantic nor permissive, what kind of information is it? The idea that has been missing from the analysis thus far is that of biological function. Neural information is instantiated, not *wherever* there are empirical correlations, but only where biological control systems have evolved to exploit correlations. When we estimate the quantity of information flowing through a neural

pathway, we are not simply re-expressing a physical quantity on a logarithmic scale. Instead, we are expressing a fact about a specific biological capacity of an organism.

This functional conception of information doesn't fit neatly into the existing philosophical taxonomies. Nevertheless, it is not entirely novel. The notion of information in signaling games, such as those described in Skyrms (2010), is also, at least in a very broad sense, functional. In a typical game, we have a communication setup that looks a lot like the XYR model: it includes a set of environmental states, a set of possible signals, and a set of receiver responses. In Section 2, I emphasized that if we want to quantify the information in a biological system, we must heed the distinction between functional and non-functional activity. Is this point sufficiently general to apply to the informational quantities in signaling games? In principle, yes. A vervet who looks up at the sky has not (yet) signaled that an eagle is near, even if there happens to be some correlation between looking up and the presence of eagles. So it would be a mistake to include "looking up" events when quantifying the information in vervet predator signaling. Despite this, there is no methodologically significant parallel between the notion of information in signaling games and that in neurophysiology. The primary reason for this is that signaling game models are not data driven. The insights they provide do not typically depend on accurate estimation of empirical magnitudes. Modelers simply stipulate that, for example, the environment is to be partitioned into three discrete states. In neurophysiology, the partitioning must be discovered rather than stipulated.

Moreover, as Cao (2012) emphasizes, the notion of information in the signaling game literature is only applicable within a game-like setting, where the notion of utility has some natural application. Definitions of information in the signaling game context are not, therefore, sufficiently general to cover all functional notions of information, and they are particularly awkward to apply to within-organism communication systems, where the notion of utility is undefined.

So what would a definition of information look like, if it was designed to capture within-organism functional phenomena? The following definition, developed by Bergstrom and Rosvall, provides a good starting point.

> An object X conveys information if the function of X is to reduce, by virtue of its sequence properties, uncertainty on the part of an agent who observes X (Bergstrom and Rosvall, 2008).

This definition is on the right track.[9] However, as Godfrey-Smith (2011) suggested in a response article, it is difficult to understand what exactly is meant by the term "agent" in the definition. The following dilemma reinforces Godfrey-Smith's worry. If the term "agent" is meant in the full-fledged sense of an autonomous, goal-driven decision maker, then it seems unlikely to apply to small neural mechanisms. If it is meant only as an abstraction, perhaps in the sense of an ideal observer of the system, then nature cannot have selected the object for its effects on the agent. In that case, the phrase "function of X" seems toothless, and the definition loses empirical content.

Bergstrom and Rosvall's appeal to agency is not merely a quirk in their presentation. The appeal to agency is seductive, and is deeply embedded in both philosophical and scientific discussions of information theory. Unfortunately, the appeal to agency threatens to undermine a kind of scientific objectivity that we should hope to preserve in any respectable neurophysiological theory. In the next section, I take up the question of objectivity, and try to reconcile it with the kind of receiver-relativity discussed in Sections 2 and 3.

4 Is neural information objective?

In their book *Memory and the Computational Brain* Gallistel and King provide an eminently clear and thorough account of what a successful theory of neural representation would have to include, and the various ways in which current theories fall short. One of the most prominent themes in the book is the warning that we are likely to overestimate the degree to which we understand how the brain works if we are not careful to insist on concrete, material interpretations of central concepts like "representation" and "computation." Nevertheless, in their discussion of information theory, which plays a central role in the account, they too fall back on the notion of agency. The authors, who deserve praise for their intellectual honesty, are explicit about this appeal, and quite candid about its less palatable consequences.

> This is an absolutely critical point about communicated information - and the subjectivity it implies is deeply unsettling. By subjectivity, we mean that the information communicated by a signal depends on the receiver's (the subject's) prior knowledge of the possibilities and

[9] Although their article is focused on genetic information, they suggest that their definition can be extended to include neural information quantities.

> their probabilities. Thus the information actually communicated is not an objective property of the signal from which the subject obtained it (Gallistel and King, 2009, p. 8)!

I agree that the implied subjectivity is deeply unsettling. It is unsettling because, if a physical magnitude depends on the perspective of an agent (or subject), it lacks a paradigm kind of scientific objectivity. In an overview of the philosophical literature on scientific objectivity, Sprenger and Reiss suggest that discussions of scientific objectivity typically begin with the following thought. There are, fundamentally, two kinds of qualities in the world... "the ones that vary with the perspective one has or takes, and the ones that remain constant through changes of perspective" (Reiss and Sprenger, 2014, p. 4). What it means to say that a body of scientific theory is objective is that it restricts itself to properties of the latter sort. If informational quantities depend on the epistemic state of an agent, then they are not invariant to shifts in perspective, and therefore lack this basic variety of scientific objectivity. That is a conceptual flaw we should be unwilling to accept.

Before I give my own definition of information, let us ask why scientists and philosophers alike so frequently rely on notions of agency to describe informational quantities. The root of the problem may be the overwhelming temptation to rely on analogies to cases of information measurement in human communication. Here is a simple case that illustrates how such analogies generally work. A politician on the campaign trail has four prepared speeches, and chooses one for each scheduled speaking event. For most of the people in the audience, the speech contains lots of information. But for the speech writer, who, let us suppose, is tagging along on the campaign trail, the quantity of information communicated is much smaller. For her, there are only four possible signals, and all the uncertainty associated with the event is resolved as soon as the first few words are uttered. If we assume that the speeches are chosen at random and with equal probability, the speech writer receives exactly two bits of information as soon as she recognizes which of the four speeches has been selected. In this situation, the epistemic state of the receiver (the speech writer) effectively partitions the source into distinct signals, and the manner of that partitioning determines the quantity of information transmitted. If we model all instances of information transmission on cases like this, it is hard to suppress the suspicion that informational quantities have an irreducibly subjective quality.

Agential definitions of informational quantities do get something right: they

highlight the relativistic nature of information. But it is not necessary to rely on the notion of agency to do this. Instead, we can simply insist that the informational quantity associated with a signal depends on the role that the signal plays in the functional economy of the organism. Notice that the political speech story bears structural similarities to the XYR model explored in Section 2. In both cases, properties of the receiver have a direct influence on the probabilities attached to the signals. In the political speech case, the influence is epistemic; it has to do with the speech writer's priors. In the XYR model, there is no agent to whom such Bayesian properties could be ascribed. The influence is instead a matter of functional capacity. This parallel suggests a way forward for constructing a definition of information that captures the concept at work in neural information estimates without relying on agency. I'll use the Bergstrom and Rosvall definition as a starting point.

> An object X conveys information if (i) the sequence properties of X are correlated with the states of some variable Y that has biological relevance to the organism and (ii) there exists a receiver mechanism R, whose function it is to exploit the correlation between X and Y to realize some biological capacity.

This definition is admittedly loose, and is not intended to serve as anything like a procedure for sorting informational systems from non-informational systems. The difficulties of that task are buried in the meaning of the terms "sequence properties" and "receiver mechanism," the interpretation of which will vary dramatically from one biological system to the next. It is even less useful for determining how much information is flowing through a biological communication system. It does, however, manage to present the notion of an informational quantity in a way that acknowledges its relativistic character without hitching it to the troublesome notion of agency.

When interpreted as claims about *this* kind of information, neural information estimates have the potential for a substantial kind of scientific objectivity. This is because the kind of relativity at issue is relativity to the capacities of a biological mechanism which we can, at least in the best cases, identify and observe. Of course, this view also suggests that neural information estimates are subject to the vagaries of functional analysis. In those cases where the functional role of a neuron is clearly specifiable, such as it is in the case of the H1 neuron, estimates of neural information are reasonably objective. In cases in which the functional role of a neuron is less clearly specifiable, the correct

experimental procedure for determining the role of noise in the system will be underdetermined by the facts at hand. Or, when a single neuron plays multiple functional roles at once which cannot easily be disentangled experimentally, we should expect that no particular estimate will be the final word one the matter.

5 Conclusion

I have emphasized that neural information estimates quantify functional capacities. Their values are entailed by underlying correlations between empirical variables. However, because they have a fundamentally functional character, they cannot be estimated accurately without taking into account the manner in which those correlations are used by downstream mechanisms. The concept of information implicit in information estimates of neural activity is novel; it doesn't correspond to either of the most prominent conceptions of information in the philosophical literature. Finally, I argued that, despite appearances to the contrary, there are cases where neural information estimates can be regarded as robustly objective properties of a neural system, despite the relativity to which they are inevitably subject.

References

Adams, F. (2003). The informational turn in philosophy. *Minds and Machines*, 13(4):471–501.

Bar-Hillel, Y. and Carnap, R. (1953). Semantic information. *The British Journal for the Philosophy of Science*, 4(14):147–157.

Bergstrom, C. T. and Rosvall, M. (2008). The transmission sense of information. *Biology and Philosophy*, 26(2):191–194.

Borst, A. and Theunissen, F. E. (1999). Information theory and neural coding. *Nature neuroscience*, 2(11):947–957.

Cao, R. (2012). A Teleosemantic Approach to Information in the Brain. *Biology and Philosophy*, 27(1):49–71.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley-Interscience, New York, 1 edition.

Dayan, P. and Abbott, L. F. (2001). *Theoretical neuroscience*, volume 10. Cambridge, MA: MIT Press.

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

Floridi, L. (2015). Semantic conceptions of information. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2015 edition.

Frye, M. A. and Dickinson, M. H. (2001). Fly flight: A model for the neural control of complex behavior. *Neuron*, 32(3):385 – 388.

Gallistel, C. R. and King, A. P. (2009). *Memory and the Computational Brain: Why Cognitive Science will Transform Neuroscience*. Wiley-Blackwell, Chichester, West Sussex, UK ; Malden, MA, 1 edition.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, 28(3):344–362.

Godfrey-Smith, P. (2011). Senders, receivers, and genetic information: comments on bergstrom and rosvall. *Biology & Philosophy*, 26(2):177–181.

Godfrey-Smith, P. and Sterelny, K. (2008). Biological Information. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition.

Harms, W. F. (2006). What is information? three concepts. *Biological Theory*, 1(3):230–242.

Magal, C., Dangles, O., Caparroy, P., and Casas, J. (2006). Hair canopy of cricket sensory system tuned to predator signals. *Journal of theoretical biology*, 241(3):459–466.

Neri, P. (2006). Spatial integration of optic flow signals in fly motion-sensitive neurons. *Journal of neurophysiology*, 95(3):1608–1619.

Piccinini, G. and Scarantino, A. (2011). Information Processing, Computation, and Cognition. *Journal of Biological Physics*, 37(1):1–38.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L., LaMantia, A.-S., McNamara, J. O., and Williams, S. (2001). *Neuroscience*, volume 3. Sinauer Associates, Sunderland, MA.

Reiss, J. and Sprenger, J. (2014). Scientific objectivity. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

Rieke, F., Warland, D., Steveninck, R. d. d. R. v., and Bialek, W. (1999). *Spikes: Exploring the Neural Code*. A Bradford Book, Cambridge, MA.

Sayre, K. (1976). *Cybernetics and the Philosophy of Mind*. Routledge.

Shannon, C. E. (1956). The bandwagon. *IRE Transactions on Information Theory*, 2(1):3.

Shannon, C. E. and Weaver, W. (1949). The mathematical theory of information.

Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford University Press.

Van Hateren, J., Kern, R., Schwerdtfeger, G., and Egelhaaf, M. (2005). Function and coding in the blowfly h1 neuron during naturalistic optic flow. *The Journal of neuroscience*, 25(17):4343–4352.