

# THREE APPROACHES TO INFORMATION AND THEIR RELEVANCE TO THE NERVOUS SYSTEM

Christopher D. Fiorillo

**Our understanding of the information processing function of the nervous system has not kept pace with our understanding of its mechanical function. I suggest that a significant reason for the lack of progress arises from the dominant philosophical and mathematical approach to information, rather than merely the complexities of the nervous system. The concepts of information and probability are dependent on one another, and yet many neuroscientists are not aware that there are sharply different definitions of probability as well as different approaches to its application. First, there has been a lack of recognition that all probabilities are conditional on information, as specified by a “Bayesian” definition. Second, whereas the conventional approach has been for scientists to describe their knowledge about the relationship between a system’s inputs and outputs (a third-person perspective), Bayesian theory provides a formal method to describe what information a system has about its own inputs and outputs (the first-person perspective). I propose that a strictly Bayesian, first person perspective on information and probability will lead to a better understanding of nervous system function and more rapid progress towards artificial intelligence.**

The nervous system is often referred to as “an information processing machine.” Although some may disapprove of its characterization as a “machine,” no one would argue against the assertion that its purpose is to process information. However, there is a lack of consensus about the meaning of information and how it can be measured. Although this lack of consensus is a particular problem within the area of theoretical and computational neuroscience, it is virtually never acknowledged within the literature that there is any meaningful disagreement. Thus many researchers within this field may not be fully aware of the distinct views or of their importance.

Three approaches to probability ought to be distinguished with respect to the study of a neural system. The system could be an ion channel, a neuron, or a nervous system, and we are interested in the probabilities of its variable inputs and outputs. Probabilities can be viewed as a property of the inputs or outputs themselves (the “frequentist” view), or as a property of a scientist’s information about the input or outputs (a “third-person” Bayesian view), or as a property of the system’s information about its own inputs or outputs (the “first-person” Bayesian view). An ill-defined mixture of these first two perspectives has dominated neuroscience. The purpose of this article is to draw awareness to the differing perspectives on information, and to argue that a first-person perspective will lead to a better understanding of neural function. The relevance of these differing approaches extends beyond formal mathematical treatments to basic qualitative models that we use to understand the nervous system.

## **Two Definitions of Probability**

It is agreed upon that information is a reduction in uncertainty (or more specifically, a measure of uncertainty known as entropy), as described by Shannon (1948). Uncertainty (entropy) refers to the width or ‘flatness’ of a probability distribution. A perfectly flat probability distribution, in which all possible states are equally probable, corresponds to complete ignorance, or an absence of any information. The greater the information, the narrower the probability distribution and the less entropy it has. However, the definition of probability has been disputed, and thus so has the effective definition of information.

Although virtually unacknowledged within neuroscience, two fundamentally distinct definitions of probability have been proposed, “Bayesian” and “frequentist.” The two definitions can be illustrated by the differing answers that might be given to a simple question: “An event has four possible outcomes, *A*, *B*, *C*, and *D*. What is the probability of each outcome?” The Bayesian answer is that, in the absence of any additional information, logic requires that all outcomes be considered equally likely, and thus the probability of each is 0.25. (“Bayesian” comes from Thomas Bayes, who in the 19th century gave the first formal mathematical description of probabilities as being conditional on information.) The frequentist response to the same question is that the question is inappropriate. In order to apply the concept of probability, one must first observe the frequency with which each outcome occurs. As one makes repeated observations of the outcome, one can begin to “estimate” the probabilities. To be certain of the probabilities, one would have to make an infinite number of observations. According to a strict frequentist definition, probabilities are equivalent to frequencies; they are a physical property of a system, and they are independent of an observer’s knowledge of the system. This is the view taken by Feller (1950) in his classic work on probability theory.

### **The Bayesian Definition of Probability**

Bayesian theory ultimately seeks to provide a unified account of information (knowledge) and reason (logic), and thus its relevance extends far beyond formal applications of probability theory. The Bayesian definition of probability has been advanced by many authors over the years, but the account given here is based upon the recent textbook of Jaynes (2003). According to this view, probability theory is a natural extension of earlier work on logic. By insisting that propositions be either true or false, conventional logic is not applicable to the conditions of uncertainty that characterize nature. Bayesian probability theory incorporates uncertainty by describing confidence (or strength of belief) in a proposition on a continuous scale of 0 to 1. The probability of a particular proposition (or state) is always entirely conditional on a particular set of knowledge or information. The rules relating knowledge to probability are essentially just “common sense” (at least for simple states of knowledge). Indeed, Laplace (1819) referred to probability theory as “nothing but common sense reduced to calculation.” Through logic, information specifies probabilities, and likewise, probabilities describe and quantify information. The conditional relationship between information and probabilities is the critical defining feature of what I refer to here as the Bayesian definition (regardless of whether or not Bayes’s theorem is formally utilized).

Much of what we mean by “common sense” is embodied mathematically in the principle of maximum entropy. The maximum entropy principle (or logic) requires that we fully acknowledge our ignorance by considering all possibilities equally probable unless we have evidence to the contrary. For example, if the only information available is that “there are four possible outcomes,” then the probability distribution that describes that information is “flat” since entropy is maximized when the probabilities are all equal. Since by definition the sum of the probabilities must equal one, the probability of each outcome is 0.25. In contrast to this contrived example, we often have information that does not constrain the number of possible outcomes, but does constrain the mean and the variance. Such a state of knowledge often corresponds to a Gaussian probability distribution, which has the maximum entropy for a given variance. In other cases our knowledge derives from observing the past frequency of an event, in which case the probability distribution that best describes our knowledge may closely resemble the observed frequency distribution. Thus knowledge derived from measurement of frequency distributions is treated just like any other knowledge.

Several criticisms have been made of Bayesian probabilities. Extensive counterarguments have been given elsewhere (e.g. Jaynes, 2003), but I will briefly summarize several important points here.

First, Bayesian probabilities have been criticized as being “subjective,” and thus ambiguous and not appropriate for science. The characterization of Bayesian probabilities as “subjective” can be misleading. According to a Bayesian definition, if two entities possess different information, then reason requires that they assign different probabilities to the same event. In this sense, probability and uncertainty are subjective and relative rather than objective or absolute properties of the world. However, a given state of knowledge uniquely determines a single probability for a particular event through the principle of maximum entropy or logic. Therefore, any two rational entities possessing exactly the same knowledge must assign the same probabilities to all events, and in this sense probabilities are objective properties of a set of information. Thus Bayesian probabilities are not in any way subjective beyond their acknowledgment of the reality that the world looks different from different vantage points. As discussed further below, this relativism is what makes Bayesian probabilities so useful in understanding brain function, since different brains obviously have different information.

A second criticism of Bayesian probabilities is that it is not always clear how they should be calculated. Although this is undoubtedly true, it is not a valid criticism of Bayesian principles. To derive Bayesian probability distributions, one must first be able to specify precisely what information is relevant. In our subjective conscious experience, and even in the conduct of science, this is often not possible as a practical matter. We often have a large amount of relevant information from various sources, and it is not explicitly quantified in any sense. In such cases, probability theory as a formal tool may be of little use. Indeed, it is largely the case of increasingly complex information states that occupies researchers in the field of probability theory over the years. But in our daily mental lives, we constantly work with large and diverse sets of information, and it would be odd to argue that this ‘cognitive’ information is fundamentally distinct or that it cannot be quantified as a matter of principle (although it may indeed not be possible as a practical matter).

A third criticism of a Bayesian approach stems from confusion of the subjective aspect of Bayesian probabilities with our conscious attempts at quantifying the strength of our beliefs. A person often struggles to state the probability that one of their beliefs is true. This may be in part for the same reason that scientists and experts on probability theory struggle to rationally calculate probabilities in cases in which a great diversity of information is relevant. However, although human behavior is routinely based upon perceptions of what is probable and what is not, to be asked to verbally state a probability is highly unnatural. The brain was designed to process information in order to generate adaptive behavior, but not, in general, to quantify its own information. In principle, Bayesian methods could be used to quantify the information in a person’s brain, as summarized further below.

### **Faults with the frequentist definition of probability**

The faults of the frequentist approach have been extensively documented elsewhere (e.g., Howson and Urbach, 1991; Jaynes, 2003). One fault that is worth mentioning here is that although frequency distributions are often measured within formal science, there are numerous cases of information for which there is no relevant frequency distribution. Perhaps the simplest example would be a statement of the type given above that specifies only the number of possible outcomes. Since there are no frequencies, this information simply cannot be the basis of a probability distribution within a frequentist framework. The problem is not merely that frequency distributions may not be available as a practical matter, but rather that in many cases no relevant frequency distribution has ever existed or ever will exist. In cases where frequency distributions do exist, it is not clear over what finite range or period they ought to be measured in order to derive probabilities. Whereas Bayesian methods could

be used to describe and quantify any information, at least in principle, frequentist methods are frequently inapplicable.

The frequentist view of probability may be very slowly falling out of favor, but it still exerts a dominating influence within neuroscience. This is illustrated well by a remark made in 2004 by a prominent neuroscientist: “How can probabilities of external events be conditional on the internal information an animal has, unless we assume telekinesis?” (Further dialogue suggested that this person did not reject the Bayesian definition of probabilities, but had merely been unaware of it.) Another vivid example of the frequentist perspective comes from a 2010 ‘Review’ in *Nature Reviews Neuroscience*, which arguably represents the most authoritative source of dogma within neuroscience. Friston (2010) attributes entropy to a fish, stating “A fish that frequently forsook water would have high entropy.” He goes on to state “A system cannot know whether its sensations are surprising...” These statements demonstrate that a frequentist view, in which probability and entropy are properties of objects rather than observers, is still pervasive. Furthermore, they explicitly forbid a first-person Bayesian view.

The influence of the frequentist view is not immediately obvious because it is routinely used without any acknowledgment that more than one definition of probability even exists. I suspect this is in part because many neuroscientists are unaware of the different definitions of probability. Whether or not they are aware, the frequentist approach serves as the default. Even books that make extensive use of probabilities and quantify information do not state a definition of probability or mention that there is more than one definition (e.g. Rieke et al., 1997; Dayan and Abbott, 2001; Trappenberg, 2002).

Although the approach to probabilities is seldom stated, one can often infer a frequentist view from the author’s choice of words. For example, it is common to find language related to “estimating” or “measuring” a probability distribution. This language makes sense only if probabilities are essentially the same as frequencies, a property of an observed system rather than a property of the observer. Furthermore, it is routine to refer to the conductance of single ion channels, or the release of neurotransmitter-containing vesicles, or the spiking of single neurons, as “random” or “noisy” or “stochastic” or “probabilistic.” The strong implication is that these are intrinsic properties of these physical systems, rather than merely a description of our own ignorance. However, it is known that these systems, even single ion channels, can exist in numerous states that cannot be directly measured and are thus “hidden” from observation. These hidden states act ‘behind the scenes’ to determine the observed output of the system. According to scientific dogma, it is only at smaller subatomic scales that physical systems are intrinsically probabilistic. One could question how physicists could ever be confident that their inability to predict the behavior of a system is not merely due to their own ignorance. But since the present topic is neuroscience rather than quantum physics, we do not need to address that issue here. In neuroscience, “random” and related words, if they must be used, should be understood to signify the uncertainty of the scientist observing the system, rather than to represent a property of the system. By contrast, the variability inherent in a frequency distribution is indeed a property of a system. Adoption of a Bayesian view would therefore bring with it a change in the language we use to describe the nervous system.

Even the use of Bayes’s theorem does not indicate a strictly Bayesian view. Bayes’s theorem describes how two sets of information ought to be integrated. These are commonly referred to as the “prior” and the “likelihood.” Whereas the likelihood is always specified to be conditional on some set of information, the prior is typically not shown to be conditional on any information. In most cases (e.g. Rieke et al., 1997), the prior is derived from a frequency distribution, but the authors do not specify whether the probabilities are a property of the frequency distribution itself, the third-person

knowledge of the scientist about the frequencies, or the first-person knowledge of the nervous system about the frequencies.

Like any other intellectual endeavor, the frequentist approach does rely on the application of reason to information that does not derive from frequency distributions, even though its advocates do not explicitly acknowledge this information or how they are using it. Thus if Bayesian theory is viewed in a broad sense as the application of reason to information, then the frequentist approach may in fact be understood as a poorly formulated and incomplete implementation of Bayesian principles. A strict frequentist definition is not actually viable, since the generation of a probability distribution always requires additional information that does not come from measured frequencies. Indeed, some scientists hold the naive viewpoint that the Bayesian versus frequentist debate is meaningless because they are merely different formalisms that result in the same quantitative results. In those cases in which probabilities are conditional on knowledge of frequency distributions, both methods do in fact often (but not always) yield the same results. Thus there is a large set of cases, including many routine statistical tests that are commonly used in biological research, for which there is not much quantitative difference between the two approaches.

So why does it matter so much which approach one takes to probabilities? The Bayesian approach is more widely applicable and in some cases yields different and superior results. By explicitly incorporating logic or reason, Bayesian probability theory encompasses a much greater domain than the narrow confines of statistics. Bayesian probabilities are fundamentally distinct from and incompatible with a strict frequentist view. But my main reason for presenting the argument for a Bayesian perspective is that once one accepts the Bayesian perspective, then one can readily understand that there are radically different Bayesian approaches that one might take towards characterizing brain function. I suggest below that the conventional third-person approach is not the most useful for understanding how the brain processes information, whereas the first-person perspective may be simpler and provide greater insight.

### **First Versus Third-Person Perspectives**

The critical feature of a Bayesian perspective, which is explicitly forbidden by a frequentist perspective, is that it allows us to describe and quantify sets of information other than our own, or that of Science. This is a great virtue, since different brains have different information. However, before exploring the application of Bayesian theory to distinct sets of information, we should first have a clear understanding of its conventional use in science. Bayesian theory describes how logic should be applied to information. But of course people, and especially scientists, were doing that long before the methods were formally written down and called ‘Bayesian theory.’

Scientists ideally try to work from a common, shared body of knowledge. To the extent that two rational scientists share the same information, they will naturally agree on the probabilities. (Indeed, the sharing of information is why the Bayesian versus frequentist debate sometimes appears to be a minor technical squabble.) The conventional approach of scientists is to describe nature from the common perspective of a unified Science, made possible by the sharing of information. In describing a system such as the brain, or a neuron, or an ion channel, scientists naturally describe the relationship between its inputs and outputs from their own perspective. The scientist is an observer of the system under study, and thus the role of the scientist is analogous to the role of a “third-person” narrator in literature, whereas the observed system (brain, neuron, or ion channel) would be the “first-person.” The first-person perspective would be that of the system observing its inputs. Here I use this “first” and “third person” distinction to describe these two perspectives, or sets of information. There are numerous third-person perspectives, corresponding to different observers, whereas the first-person

perspective is unique. But here I refer to the ‘third-person perspective’ to mean that of a scientist observing a neural system.

The conventional approach of the scientist is to view the system as the third-person observer, measuring and manipulating its inputs and outputs to map the “input-output” (I-O) relationship. An extreme and famous example of this approach is provided by Skinner’s work on animal behavior. Skinner has been criticized for his emphasis on the study of inputs and outputs without reference to the “black box” (brain) that lies between. However, throughout science it is standard practice to characterize I-O transformations with little regard to the black box that performs the transformation. A scientist much by necessity choose a level of analysis and then ignore other levels. Clearly the study of I-O relationships has been very beneficial in neuroscience as it has throughout science. The problem in neuroscience is that the I-O relationships are often exceedingly complex and therefore attempts to characterize them often provide us with very limited insights. Even single ion channels exhibit complex behavior, with numerous unobserved states that act “behind the scenes” and make it difficult for us to predict when a channel will be open or closed. Similarly, Skinner and other scientists have often taken a third-person approach to predicting a persons outputs (behaviors) by observing inputs and trying to characterize the I-O relationship.

The third-person perspective is undoubtedly useful, but a first-person perspective may be simpler and provide us with greater insight. Indeed, outside of formal science, people almost always utilize a first-person approach. To predict the actions of someone else, one tries to “see the world through their eyes” by inferring their state of knowledge. To the extent that one person knows what information another person possesses, that person’s behavior can often be accurately predicted. In psychology, a person’s understanding that other people possess different information about the world is called a “theory of mind.” A theory of mind appears to be lacking in infants and in most animals, and it represents a tremendous developmental and evolutionary advance in cognition. Scientists obviously possess a theory of mind, but they have seldom put it into practice by taking the first-person perspective (perhaps out of a misguided fear of losing objectivity). Bayesian theory lays out the principles by which we could attempt to give an objective, mathematical description of the knowledge of an individual person, or neuron. Adoption of a first-person Bayesian perspective may therefore advance neuroscience in much the same way that a theory of mind allows each of us to better understand and predict the actions of others.

The simplest way for us to predict a person’s behavior is to question that person about what he or she knows about the world, including his or her plans. Extracting first-person information from an animal’s brain (or neuron or ion channel) is not so easy, but we do know a great deal about the mechanics of these systems. This knowledge, together with Bayesian principles, could allow us to describe and quantify a system’s information about its world. Below I discuss how this might be done.

### **The First-Person Bayesian View Applied to Perception and Behavior**

Hermann von Helmholtz argued in the 19<sup>th</sup> century that the nervous system must infer or estimate the structure of its environment. This description of nervous system function is clearly from the first-person perspective of the nervous system. Although there has been widespread agreement with von Helmholtz’s assessment, progress has been limited by the persistent, but unacknowledged and perhaps inadvertent, influence of third-person and frequentist perspectives.

A good starting point for choosing between first- and third-person approaches is to consider which one provides a simpler description and greater insight with respect to nervous system function. In either case it is clear that the goal of the nervous system is to select outputs so as to promote the biological fitness of the animal. But from a third-person perspective, it is not at all clear how the

system could achieve this goal. Obviously the system needs to identify and select the most appropriate I-O transformation for each context, but this fact provides little insight.

From a first-person perspective, the selection of the most appropriate outputs can be understood as decision-making, something with which we all have relevant experience and intuitions. The problem in decision-making is uncertainty, or lack of information. If I am certain about the state of the world and how the world works, then the problem is solved and I simply select the output that I know will yield the best future outcome. Of course absolute certainty is out of reach in this world, but the nervous system should minimize its uncertainty. This is essentially the same function identified by von Helmholtz, since an accurate inference or estimate of the state of the world is one with low uncertainty. Likewise it fits with our understanding that the nervous system is an ‘information processing machine,’ since minimizing uncertainty is the same as integrating information. I recently proposed a general theory of nervous system function that starts with the assumption that the nervous system seeks to minimize its uncertainty about reward-related aspects of the world (Fiorillo, 2008). The critical point to the present argument is that minimizing uncertainty is obviously not the goal of the system from the third-person perspective, since the uncertainty of a third-person observer is not relevant to the function of the system.

It is natural to adopt a first-person perspective in studying human perception, and in the last decade substantial progress has been made in this endeavor through application of Bayes’s Theorem, which describes how two sets of information ought to be integrated to make an inference or estimate (Knill and Richards, 1996; Weiss et al., 2002; Rao et al., 2002; Yang and Purves, 2003; Singh and Scott, 2003; Purves and Lotto, 2003; Niemeier et al., 2003; Kording and Wolpert, 2004; Knill and Pouget, 2004). In applying Bayes’s Theorem to the nervous system, one set of information is ‘prior,’ typically meaning it is information about the external world that is already in the system (as a result of learning, whether over the last few seconds or over evolutionary timescales). The other information is the incoming sensory ‘evidence.’

Some of the best evidence that the brain rationally integrates sensory with prior information comes from studies that have used Bayes’s Theorem to explain what may have appeared to be serious flaws in brain function. There are many documented examples in which people consistently misperceive the external world. However, illusions and other misperceptions may be explained as rational (optimal) inferences based on limited information, and they often occur when the brain is presented with sensory patterns that are unusual in a statistical sense. For example, the direction of movement of a drifting grating is ambiguous when viewed through an aperture. The sensory evidence is consistent with any direction of movement across a range of 180 degrees, but movement in some directions would require higher speed than others. The single direction that is perceived is that which is consistent with the slowest movement, since objects in the natural world tend to move more slowly if at all, whereas fast motion is unusual (Weiss et al., 2002). Whereas that study and others assumed that the brain has prior knowledge of the world’s ‘natural’ statistics, Kording and Wolpert (2004) trained subjects in a sensorimotor task and thereby demonstrated that the prior information used by motor systems is not static but is subject to learning.

Whereas some studies have utilized Bayes’s Theorem in a formal sense, others have applied the same types of principles in a less quantitative form. For example, linguistic abilities may be understood as a probabilistic inference problem (Seidenberg, 1997). In ‘binocular disparity’ experiments, if a person is presented with a distinct but similar scene in each eye, the person perceives the average of the two scenes if the average is the sort of scene that is consistent with the natural world. But when the two scenes are quite different and cannot be combined in a manner that is consistent with what is known about the natural world, then the person perceives one or the other scene

but not both. In the McGurk effect, a person's auditory system experiences one sound while watching someone moving lips in a manner that is associated with another sound. The person's perception of the spoken word corresponds neither to the auditory nor visual information but rather to an 'average' of the two, presumably because this average best explains the sensory evidence given the prior information. Similarly, one can imagine how rational inference could explain the placebo effect, since the trusted advice of a medical authority, or any other cause for hope, is at least a bit of evidence that things are not as bad as the sensory evidence from an injured body part would otherwise lead one to believe.

### **The Conventional Third-Person Approach to Neurons**

If the first-person approach provides a simpler description of nervous system function and accounts better for perception and behavior, then one would expect that it is also the better approach for understanding how neurons contribute to perception and behavior. However, the actual approach taken to studying the nervous system has depended strongly on the level of analysis. Ion channels, and neurons, and even nervous systems of simple animals, are routinely viewed from the third-person perspective as 'objects' to be understood by characterizing their I-O relationship. By contrast, within psychology and psychiatry it has been natural to take the first-person perspective of the 'subject,' with the notable exception of Skinner and behaviorism.

The first-person perspective has likewise been extended to understand those regions of the brain that are most 'cognitive,' that lie between the sensory and motor peripheries and are most closely related to consciousness. An example that is related to my own research comes from experiments on dopamine neurons of the ventral midbrain in monkeys (Fiorillo et al., 2003; Tobler et al., 2005). The activity of these neurons depends on predictions of reward value, and the predictions change dynamically as the animal learns through experience. The predictions are best thought of in probabilistic terms, and it seems obvious to almost everyone that the predictions and probabilities are entirely conditional on the information possessed by the animal. But if one were to do an analogous experiment with a neuron in culture, or in a fly's brain (e.g. Fairhall et al., 2001), or in a sensory structure such as the retina, then the standard approach is to take the third-person perspective and to characterize the neuron's I-O relationship. The change in neuronal responses might be explained as an "adaptive filter" rather than as a system that "learns to predict." The fly neuron and the retinal neuron are described as objects from the perspective of the scientific observer, whereas the monkey neuron is described from the perspective of the monkey, or subject. It is apparent that for 'higher' functions of the nervous system we apply a theory of mind, but for 'lower' functions we do not. Although this dualistic approach is entrenched within neuroscience, I am unaware of any arguments that have been made in its defense. Obviously this schism makes it difficult for the field to converge towards a more unified understanding of neural function.

The most extensive literature with respect to information and neurons concerns the "efficient coding hypothesis" proposed 50 years ago (Attneave, 1954; Barlow, 1961). This has been one of the most successful theories within neuroscience, accounting for numerous features of early sensory systems (e.g. Srinivasan et al., 1982; de Ruyter van Steveninck and Laughlin, 1996; Rieke et al., 1997; Fairhall et al., 2001; Simoncelli and Olshausen, 2001; Hosoya et al., 2005). The ideas originated with Shannon's foundational work on "information theory" (1948). Shannon, working for a telephone company, was concerned with how an engineer should design a communication channel to be efficient in transmitting information. Much of the work on information theory does not address the definition of probability, but often implies a frequentist definition and yet works by default from a third-person perspective (the frequentist definition being untenable and insufficient on its own). For Shannon's

purposes, the definition was not a practical concern, since he was justified in assuming that the sender, the receiver, and the engineer all share approximately the same knowledge of language.

This third-person 'engineering' approach was transferred to studies of sensory systems. But whereas sender, receiver and observer share the same basic knowledge in Shannon's engineering scenario, the same does not hold true when the sender and receiver are neurons observed by a scientist. It is a legitimate exercise to quantify information entirely from the third-person perspective, but the scientist should not confuse his or her perspective with that of a neuron.

I will focus here on the work of Rieke and colleagues (1997), since their book represents perhaps the most comprehensive effort to date to relate information to neural function in a quantitative sense. Although they give an excellent rationale for why it is desirable to "take the neuron's point of view," their analysis is actually from a mixture of frequentist and third-person perspectives.

At each moment in time, a neuron receives information through its synaptic inputs about the intensity of an external stimulus. By characterizing the variable I-O relationship of a neuron and then observing that neuron's output at a particular moment in time, Rieke and colleagues derive a probability distribution of potential input (stimulus) intensities given observation of the neuron's output. The fact that a neuron's output (firing rate) varies even when the input is constant is critical to this analysis, since this "noise" determines the uncertainty in the estimate. But it is well known that the cause of this variability is the numerous states of the neuron that are 'hidden' from the observation of the scientist in this type of experiment. For example, at one moment the neuron is hyperpolarized due to a high potassium conductance, but at a later moment it is not, even though the stimulus intensity is the same in both cases. Thus the two instances are not equivalent from the neuron's perspective, and likewise the neuron's output distinguishes the two. But from the experimenter's third person perspective, the hidden state of the potassium conductance has contributed noise and degraded the information in the neuron's output. To see the world from the neuron's first-person perspective, we would first need to know the state of the neuron, not merely the frequency distributions of its inputs and outputs.

The other information that Rieke and colleagues incorporate is 'prior' information. But this is simply the frequency distribution of the stimulus intensities, which they control as the experimenters. In a footnote on page 23, they seem to suggest that by controlling the frequency distribution of inputs, they have avoided the controversy over the definition of probability. They do not even argue or speculate that the neuron possesses any information about the frequency distribution.

The analysis of Rieke and colleagues suffers from what Jaynes referred to as "the mind projection fallacy," mistaking the scientist's ignorance for the neuron's ignorance. Information is rigorously analyzed without any distinction made between the neuron's information and the information of those doing the analysis. Thus their analysis lacks a theory of mind. Since they explicitly express their desire to "take the neuron's point of view," then it would be natural to infer that they are either unaware of the first-person Bayesian approach presented here, or that they reject it on other grounds, such as a commitment to a frequentist view. Since they never specified a definition of probability, we cannot be sure of their reasoning.

Other quantitative work on neuronal information has taken the same basic approach as Rieke and colleagues (e.g. Seung and Sompolinsky, 1993; Sanger, 1996; de Ruyter van Steveninck and Laughlin, 1996; Pouget et al., 2000; Dayan and Abbott, 2001; Fairhall et al., 2001; Simoncelli and Olshausen, 2001; Knill and Pouget, 2004; Deneve, 2008), but without necessarily claiming to take the neuron's perspective. In Bayesian terms, 'likelihood functions' have typically been shaped by the experimenter's knowledge of variability in a neuron's output ("noise") without any account of its 'hidden' states. In those cases in which 'priors' have been used, they have been determined by the

experimenter's knowledge of a frequency distribution and they have had no neuronal substrate. Thus these analyses are from the third-person perspective.

The purpose here is not to argue that the third-person perspective is invalid or without utility. However, I wish to focus on two criticisms of this third-person approach. First, the authors do not explicitly state that they are taking the third-person perspective, and in some cases they imply that their perspective is the same as the neuron's perspective. But as discussed above, the two perspectives (sets of information) are not the same. Thus we can conclude that the numbers they derive are only from their perspective, and as a third-person perspective, their numbers are not unique. For example, if a second experimenter were to repeat the same quantitative analysis, but were to use a different experimental technique that revealed a state of the neuron (such as potassium conductance) that had been 'hidden' from the technique of the first experimenter, then the two experimenters would incorporate different sets of information and would naturally derive different probabilities and quantities of information.

The second point is simply that the third-person perspective is quite complex, and a first-person perspective may be simpler. Rieke and colleagues (1997) make the same argument, and they then go on to inadvertently provide evidence for the complexity of the third-person approach by filling a book with equations that most neuroscientists would find quite unnerving (despite their simplification of Gaussian priors). The suggestion is that a neuron's perspective may be simpler because a neuron may have access to less diverse forms of information than the scientist, and because the neuron's output must be successfully interpreted by its postsynaptic neurons, if not by scientists.

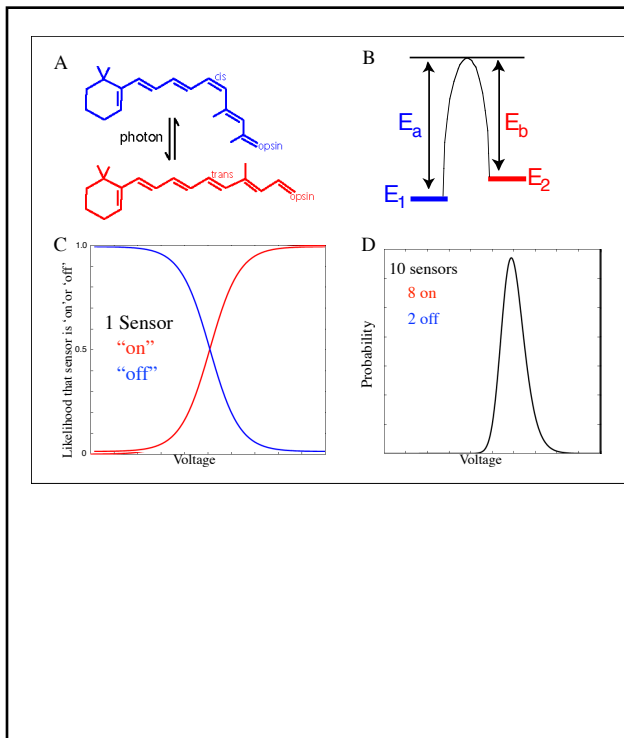
### **The First-Person Approach to Ion Channels and Neurons**

I recently published a paper that had the ambitious goal of laying out a general computational theory of the nervous system (Fiorillo, 2008). According to the theory, the computational goal of the system is to minimize uncertainty about behaviorally relevant aspects of the world, as briefly described above. The main focus of the theory was concerned with how neural plasticity can allow neurons to learn the statistical structure of their environments, and how the system can take information about relatively uninteresting quantities such as light intensity and extract information relevant to reward and decision-making. But an important part of the theory was to explain exactly what it means to say that the system "minimizes uncertainty" and to demonstrate how this occurs.

Information comes to a neuron through sensors that are coupled to ion channels. Rhodopsin is well known to be the light sensitive molecule of some photoreceptors (Figure 1A), and in a typical neuron, an analogous role is played by neurotransmitter receptors. For example, a glutamate sensor is coupled to an ion channel, all within one molecular complex. Real ion channels assume multiple configurations or states, but we can consider the simple case of a two-state sensor. Consideration of a two-state sensor is an appropriate place to start since its properties have been thoroughly described, and since more complex entities, such as an ion channel or a neuron, could in principle be modeled as a system of interacting two-state sensors. A two state sensor is the simplest possible substrate of information.

The information of a sensor could be investigated experimentally by examining its I-O relationship. However, in the case of a two-state sensor, it can also be derived from first principles using the Boltzmann distribution, which specifies the probabilities of the two states given their relative energies. These energies are dependent on the quantity that is sensed, such as voltage or ligand concentration, which I will refer to here generically as "stimulus intensity." The information possessed by a two-state sensor can be split into two components. First, it knows how the likelihoods of the two states vary depending on stimulus intensity. This is an inherent property of the sensor that does not

vary over time (equations 1-3, see methods) and illustrated by the energy diagram of figure 1B. Second, the sensor knows its current state (but not past states). Thus, given this information, the sensor can guess whether the stimulus intensity is likely to be high or low (figure 1C). The estimate is quite crude, but this is just one sensor. Some neurotransmitter receptors have more than one sensor (ligand binding site), and a neuron has many receptors. Figure 1D illustrates the estimate made by 10 sensors. A Bayesian analysis can be performed, but in this case it may not provide any additional insight. For additional details, see Fiorillo (2008).



**Figure 1.** The information in two-state sensors. **A.** The structure of retinal, a photosensor that is part of rhodopsin. Absorption of a photon changes its conformation from 'cis' to 'trans.' **B.** An energy diagram of the two states. The difference in energy levels determines the amount of time, or the likelihood, that a sensor will be in the 'off' state or 'on' state. **C.** The energy levels are sensitive to a quantity such as voltage, and thus the likelihood of a state depends on that quantity, as described by the Boltzmann equation (equation 1). A single voltage sensor can therefore estimate voltage. **D.** The estimate of voltage made by 10 voltage sensors, 8 'on' and 2 'off.' Each sensor had the same properties as shown in panel C. It was assumed that the function of each sensor was *physically* independent of other sensors, and thus the functions shown in C for a single sensor were simply multiplied together. Statistical dependence is irrelevant here, unless there is reason to believe that the neuron itself has information about the statistical dependence, as it would if there were a physical dependence between the function of the sensors.

The simple point of this analysis is that these sensors reduce uncertainty. In the absence of any sensors, there would be no information and the probability of each possible stimulus intensity would be equal (from the maximum entropy principle). The distributions of figures 1B and 1C has lower entropy than the flat distribution, and in this sense uncertainty has been reduced. This is a rather trivial accomplishment in some sense, but it quantifies the information of real neurons, and it therefore addresses the general function of the nervous system. The difficult problem the nervous system accomplishes is to take information about some relatively unimportant sensory quantity such as light intensity and to extract the information about 'reward' that is needed to guide motor output.

What about 'prior' information? It is clear that the nervous system has and uses a tremendous amount of information that is already in the system at the moment that new sensory evidence arrives. Likewise, a neuron has many ion channels that are not directly sensitive to incoming sensory information. For example, photoreceptors and all other neurons have voltage-sensitive channels that strongly contribute to their output. The information contributed by these other channels, or inputs, has been described as prior information (Fiorillo, 2008). There is strong evidence that in early sensory neurons such as those in the retina, these other inputs function to predict and cancel the effect of excitatory 'sensory' inputs (Srinivasan et al., 1982; Hosoya et al., 2005). Inhibitory synaptic inputs contribute prior spatial information, using information about light intensities in the surround to predict those in the center. Voltage-regulated potassium channels contribute prior temporal information

concerning past light intensities. Prior information is essentially subtracted from current sensory input to generate an output that can be understood as a prediction error. This sort of error signal is widespread, occurring in neurons throughout much of the nervous system.

Precisely how this cellular prior information contributes to the perceptual and behavioral phenomena that have been studied using Bayesian methods is unclear. But it is well established that this sort of “predictive coding” is an efficient way to process information. Because of its use of prior information, the neuron’s output signals only what is new in its environment, and this new information is exactly what the system needs to improve its predictions. This interpretation is substantially the same as that of the efficient coding hypothesis. The success of the efficient coding hypothesis is likely because the neuron and the scientist have both observed the same frequency distribution and thus do indeed share *some* information with one another (although this is only true for early sensory neurons, since scientists cannot readily have comparable knowledge with respect to the inputs of higher sensory neurons).

So what has been gained through clarifying the different perspectives on information? First, we have seen that the numbers are different from a first-person perspective, and perhaps simpler to derive and understand. We have also seen that there is no unique third-person perspective, and thus the numbers derived with such methods should be interpreted cautiously. But the big advantage of the first-person perspective is that we can relate the information held within molecules and single neurons, especially in early sensory systems, to the general goal and function of the nervous system, which is to make predictions. Single neurons and early sensory systems are obviously much better understood than networks of neurons in the deep recesses of the brain. By contrast, the conventional third-person perspective of the efficient coding literature could not claim much more than having found an efficient means for packaging information in early sensory regions and shipping it off to higher areas, where presumably the hard work begins, whatever that may be. By relating single neurons and early sensory areas to the general function of the brain and to higher-order areas, the first-person perspective may contribute to a unified understanding of the nervous system.

## Methods

According to the Boltzmann equation, the probability  $P_2$  that a sensor is in state 2 depends on the energy difference ( $E_2 - E_1$ ) between the two states,

$$P_2 = \frac{1}{1 + \exp\left(\frac{E_2 - E_1}{k_B T}\right)} \quad (1)$$

where  $k_B$  is Boltzmann’s constant, and  $T$  is temperature in Kelvin. Since this is a sensor, the energy difference depends, by definition, on interaction with a quantity such as voltage or the binding of a ligand. This equation is routinely used to model the effect of membrane voltage on charged, voltage-sensitive domains of ion channels. For a voltage sensor, the energy difference between the two states in the Boltzmann equation (equation 1) is

$$E_2 - E_1 = ze(V_{1/2} - V) \quad (2)$$

where  $z$  is the number of equivalent elementary charges,  $e$  is the elementary charge in coulombs,  $V$  is voltage, and  $V_{1/2}$  is the voltage required to counterbalance the inherent energy difference between the two states so that they are equally probable. The sensor for a chemical ligand works in a similar but slightly different manner. Binding of a ligand acts to stabilize state 2 of the sensor. However, unlike the dependence of a sensor’s energy states on voltage, the relative energies of the bound and unbound

states are independent of ligand concentration. The exponential term in equation 1 is thus a constant (for a given temperature), and the likelihood ( $P_2$ ) of a receptor being bound turns out to be

$$P_2 = \frac{[L]}{[L] + K_D} \quad (3)$$

where  $[L]$  is ligand concentration and  $K_D$  is the equilibrium dissociation constant. Equations 1-3 specify the likelihood that a single ligand or voltage sensor is in the “on” or “off” conformation as a function of stimulus intensity, as shown in figure 1C.

## SUPPORTING REFERENCES

Attneave, F. (1954) Some informational aspects of visual perception. *Psychol Rev* 61: 183-193.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: *Sensory Communication* (Rosenblith WA, ed) pp 217-234. Cambridge, MA: MIT Press.

Dayan P, Abbott LF (2001) *Theoretical Neuroscience*. Cambridge, MA: MIT Press.

Deneve, S. (2008) Bayesian spiking neurons I: inference. *Neural Computation* 20: 91-117.

Fairhall, A.L., Lewen, G.D., Bialek, W., & de Ruyter van Steveninck, R.R. Efficiency and ambiguity in an adaptive neural code. *Nature* 412, 787 - 792 (2001).

de Ruyter van Steveninck RR, Laughlin SB (1996) The rate of information transfer at graded potential synapses. *Nature* 379: 642-645.

Feller W (1950) *An Introduction to Probability Theory and its Applications*. New York: Wiley

Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898 - 1902 (2003).

Friston, K. (2010) The free energy principle: a unified brain theory? *Nat Rev Neurosci* 11: 127-138.

Hosoya T, Baccus SA, Meister M (2005) Dynamic predictive coding by the retina. *Nature* 436: 71–77.

Howson, C., Urbach, P. (1991) Bayesian reasoning in science. *Nature* 350, 371-374.

Jaynes ET (2003) *Probability Theory: The Logic of Science*. Cambridge, England: Cambridge University Press.

Juusola M, French AS (1997). The efficiency of sensory information coding by mechanoreceptor neurons. *Neuron* 18: 959-968.

Knill DC, Richards RW (Editors) (1996) *Perception as Bayesian Inference*. Cambridge, England, Cambridge University Press.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27: 712-719.

Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. *Nature* 427: 244–247.

Laplace PS (1819) *Essai Philosophique sur les Probabilités*. Paris: Courcier Imprimeur.

Niemeier M, Crawford JD, Tweed D (2003) Optimal transsaccadic integration explains distorted spatial perception. *Nature* 422: 76-80.

Pouget, A., Dayan, P., Zemel, R. (2000) Information processing with population codes. *Nat Rev Neurosci* 1: 125-132.

Purves D, Lotto RB (2003) *Why We See What We Do: An Empirical Theory of Vision*. Sunderland MA: Sinauer Assoc.

Rao RPN, Olshausen BA, Lewicki, MS (Editors) (2002) *Probabilistic models of the brain: perception and neural function*. Cambridge, MA: MIT Press.

Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press.

Sanger, T.D. (1996) Probability density estimation for the interpretation of neural population codes. *J Neurophysiol* 76: 2790-2793.

Seung, H.S., Sompolinsky, H. (1993) Simple models for reading neural population codes. *Proc Natl Acad Sci USA* 90: 10749-10753.

Seidenberg MS (1997) Language acquisition and use: learning and applying probabilistic constraints. *Science* 275: 1599–1603.

Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27: 379-423.

Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Ann Rev Neurosci* 24: 1193-1216.

Singh K, Scott SH (2003) A motor learning strategy reflects neural circuitry for limb control. *Nat Neurosci* 6: 399-403.

Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc Roy Soc Lond B* 126: 427-459.

Tobler, P.N., Fiorillo, C.D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642 – 1645 (2005).

Trappenberg, T.P. (2002) *Fundamentals of Computational Neuroscience*. Oxford University Press, Oxford, United Kingdom.

von Helmholtz H (1896) Concerning the perceptions in general. In: *Treatise on Physiological Optics*. Reprinted in *Visual Perception*, Yantis S, editor. Philadelphia: Psychology Press, 2001, pp. 24 - 44.

Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. *Nat Neurosci* 5: 598-604.

Yang Z, Purves D (2003) A statistical explanation of visual space. *Nat Neurosci* 6: 632-640.