

## Neural Representation/Coding

The activity of neurons ultimately relates to the world outside the nervous system. The perceptions, planned actions, and beliefs of an animal throughout its life are reflected in this neural code. The nature of this correspondence allows animals to act and react in ways which machines cannot come close by being adaptable, robust, flexible, and fast. This correspondence between the outside world and inner neural code also provides a way of testing these neural responses by relating them to stimuli and actions in experiments. This entry will review the triumphs and problems in interpreting this neural code, the wide array of experimental data, the technical and theoretical limitations, and ultimately how this information can be distilled. In order to ground the concepts covered with specific examples, this entry will focus primarily on one particular area of the mammalian brain, primary visual cortex (V1), with some references to neural coding in the retina. Multiple levels of analysis will be considered; for example, neural coding can be approached from a low-level by noting how neural spikes relate to specific stimuli, like patterns of light in a visual image, or from a high level by understanding why you would want a neural code with those properties. The collection of analysis techniques will provide insight to help bring the vast amount of experimental data into a more unified, coherent framework.

### The Basics of Neural Coding

Neurons communicate throughout the brain using spikes of activity, known as

action potentials, on the order of 1 millisecond in duration. These spikes are caused by a transient change in ion concentrations across the cell membrane. This leads to a change in membrane voltage that can be picked up by electrodes inserted near the neuron.

Although there are methods of measuring neural response collectively (such as fMRI, EEG, optical imaging) single-unit electrophysiology offers an accurate means of measuring individual neural response both in time and space. For example, imagine trying to understand human speech by listening to a crowd of voices. EEG measures electrical impulses from the scalp. It is temporally precise but averages over space - like a microphone above a crowd. The analogy for fMRI, which measures changes in blood flow to a particular area following neural response, would be moving the microphone closer to a smaller group, but giving the average sound level every second. Single unit electrophysiology, in contrast to these other techniques, places a good microphone next to a particular speaker. It may not tell you everything about the crowd (or even the other half of the conversation) but it gives you detailed information about the individual person/neuron.

A full understanding of neural representation involves relating this response to the outside world. Early studies by Stephen Kuffler and others demonstrated that neurons in visual areas have receptive fields: regions of visual space where patterns of light can influence the firing of the neuron. The receptive fields of retinal ganglion neurons (whose axons make up the optic nerve) typically demonstrate a center-surround organization where a central region might excite (or suppress) a neuron, while a surrounding region would do the opposite - these neurons respond best to differences in light levels. For more details of these receptive fields, see the entry “Visual processing: retina”.

Later work by David Hubel and Torsten Wiesel measured the receptive fields of V1 neurons, which have significantly different responses to stimuli. These neurons fire more strongly to particular orientations of lines, along with other stimulus features. One class of neurons (simple cells) was shown to have oriented regions of alternating lighting preference. In contrast to retinal ganglion neurons, which have a circularly-symmetric, center-surround structure, these V1 cells find the difference between two or more nearby elongated regions. Later work mapped these receptive fields and showed how the particular pattern of light and dark preferences could be fit by a particular mathematical function, a 2D gabor function - the details of which will not be discussed here. A simple cell's selectivity for particular stimulus features, such as line orientation and position, are evident from a map of its subregions, or equivalently, the mathematical parameters of the gabor function representation. It is well understood, however, that such models are idealizations that account for only a portion of the neural response in these cells, and later visual stages are more difficult to characterize.

### Characterizing Neural Responses

Using spots, bars, and gratings as stimuli is helpful, as they can be fully described by a small set of numbers - such as position, orientation, and size – but often these stimuli do not provide enough variation in examples to fully probe how stimuli can affect the behavior of the neuron. Unstructured noise stimuli (like the “snow” on an old TV set that wasn't tuned properly) can also be presented to a neuron and related to the resulting neural response. For example, if all the random stimuli that produced a neural spike are

collected and then averaged together, the result is called the spike-triggered average (STA). This produces a receptive field as the simplest method of reverse correlation (see the entry “reverse correlation” for more details). However, such models require a great deal of data and have only a limited ability to characterize a neuron’s response; because of this, it is clear that the response to a set of simple stimuli does not provide a straightforward prediction of how the neuron will respond to more complex stimuli.

There is a general tradeoff when characterizing the stimuli that affect a neuron’s response. On one hand, structured, artificial stimuli (dots, bars, gratings) make implicit assumptions about neural response; they are often designed to elicit strong responses and require less data for analysis, but ultimately are less likely to generalize to new stimuli. Techniques that use random noise stimuli (reverse correlation) can make fewer assumptions by presenting a theoretically unlimited stimulus set, but often get weaker neural responses and can require more data than is possible to collect. Unfortunately, the tradeoff is ultimately a statistical one between restricting the stimuli or requiring massive amounts of data. One of the promising solutions to this tradeoff is the choice of natural stimuli. For example, in vision the stimuli can be randomly sampled portions of natural images. This limits the size of the data set substantially (in comparison to noise stimuli) while probing the space more fully (in comparison to dots, bars, and gratings). Also, visual systems have evolved to respond to natural scenes; so many of the responses may be weak without the appropriate choice of stimuli. Samples from natural images are also not as well controlled or defined as artificial stimuli - making their use in many experimental paradigms less palatable. However, they do provide a distinct statistical advantage when a full characterization is necessary.

## Efficient Coding: a high-level ecological approach to neural coding

As noted above, a number of methods are used in an attempt to characterize ‘what’ causes a neuron to respond. Such methods can give us succinct mathematical descriptions which offer some predictive value for individual neurons. However, as the mathematical models become more complex, it becomes more difficult to understand the behavior of the neuron in a coherent way. Even if we could fully describe and predict the response behavior of a neuron, the question remains: ‘why’ does the neuron respond in that particular way? For our visual example, why do V1 simple cell receptive fields have the particular pattern of light and dark preferences (2D gabor functions)?

The ecological, efficient coding approach states that the goal of sensory processing is to efficiently represent the information that is behaviorally relevant to the animal. In the case of V1, the incoming visual information is coming from the natural world. There are particular properties of natural images that would suggest some codes are better than others. For example, light intensity often only changes at contours, so encoding primarily those changes by responding to lines/edges would be more efficient. What occurs in the left eye correlates with what occurs in the right eye, so receptive fields in each eye should be related for a particular neuron. In general, natural images are highly redundant and removing these forms of redundancy in natural images would allow animals to use the information efficiently. Work by a variety of researchers, as will be discussed next, has demonstrated that many goals of the early visual system directly relate the behavior of these neurons to the mathematical properties of natural images.

Our environment is highly predictable - neighboring pixels in natural scenes are highly correlated. One first step in efficient processing is to remove such redundant information. For example, Joseph Atick and colleagues proposed that individual ganglion cells remove these correlations so that the maximum amount of information can be passed through a constraining bottleneck – the optic nerve. However, when the visual signal reaches V1 there are more than 100 times as many neurons in primate V1 as axons in the optic nerve. Clearly the signal is not required to be compressed at this stage. What other type of efficient coding is useful at this stage?

One can place coding strategies along a spectrum from local, “grandmother cell” codes to distributed codes. A local code uses one neuron or relatively few neurons to represent a single, relevant piece of information. The traditional example is a “grandmother cell” code where the firing of one particular neuron represents information, such as whether or not your grandmother is present. Such neural responses would be easy to learn from and react to (approach or avoidance, for example), but clearly this code has disadvantages. For example, there are not enough neurons in the brain to represent every potential combination of visual features. On the other extreme, a distributed code uses many neurons to represent a single, relevant piece of information. For example, compression strategies can often result in highly distributed codes since part of the goal is to fully utilize the response range of every neuron. Taken to the extreme, a fully distributed code would be unreasonable in the brain as learning from and decoding such representations can be cumbersome. It would be difficult to respond to a neuron’s firing if you need to sample input from every other neuron to interpret what that response means. Of course, depending on the input, a code can be local in some cases and

distributed in others. For example, a pixel representation is local for representing small dots, but distributed for almost anything else. What was found to be the case in V1, and also in many other areas of the brain, is that the neural code predictably falls between these two extremes for natural stimuli.

David Field, Bruno Olshausen and colleagues have shown that the principle properties of V1 simple cell responses can be explained by a strategy known as sparse coding where only a few neurons fire strongly and the rest fire weakly or not at all. In one study, a neural network was used to search for the sparsest representation capable of representing natural images. Simulated neurons derived from such an approach mathematically resembled the receptive fields of real V1 neurons. Tony Bell and Terrence Sejnowski applied a similar efficient coding goal to natural scene data by increase the amount of independence between neural filters using a technique called independent components analysis (ICA). Although a thorough explanation is beyond the scope of this article, sparse coding and ICA actually use quite similar approaches, and their respective algorithms can be related mathematically. This is one case where multiple, high-level objectives may be working simultaneously with similar results, guiding the evolution and adaptation of the animal visual system.

The efficient coding approach argues that the ultimate goal of any neural representation is to be useful ecologically. The type of representation should increase the animal's evolutionary fitness. However, more explicit and immediate objectives such as compression, sparse coding, or independent coding can appear somewhat removed from this ultimate goal without explanation. Compression in the retina is beneficial given the limited size of the optic nerve. Such a strategy allows more data to be sent with fewer

axons. The V1 coding objectives are efficient representations because they make the relevant information in the images explicit. That is, neurons are representing the relevant sparse or independent causes of the data, and with such explicit representations, it can make learning new associations simple. Other ecological reasons can exist for these objectives. It has also been argued that approaches like sparse coding are more metabolically efficient. With the correct constraints, a network of spiking neurons with a relatively small number of active neurons will use less energy than a network in which multiple neurons fire at intermediate rates. Explicit representation and metabolic efficiency are ecological justifications for the specific strategies mentioned. Whatever intermediate coding objective is considered, it must ultimately relate to ecologically relevant objectives.

More to the Story: rate/temporal codes and population codes

Neural codes are more complicated than the simple individual neuron firing rate picture alluded to above. For example, there is an ongoing debate about the extent to which the neural code is a rate code (only the average firing rate of a neuron matters) or a temporal code (the relative timing of the neural spikes is also important). These two strategies actually represent two endpoints on a continuum for analysis. If the window of time used for averaging is large, one is effectively assuming a rate code. However, if the time window decreases (say on the order of 1 ms or less) then only individual spikes are counted, giving a binary representation of a neural spike train. It is necessarily the case that more information is represented in finer windows, but interpreting the neural



response becomes more difficult. The experimenter often must decide between the need for a great deal of data to analyze timing events and the pooling or binning of data needed to gain statistical power. What is clear is that many neurons are providing more information regarding the stimulus than that described by the firing rate.

In order to even establish a rate code, experimenters often average a neuron's response to a repeated stimulus, but in many situations the organism does not have the opportunity to perform such averaging. In many cases, responses of a population of neurons can be collected, and an immediate decision can be made based on those responses. However, as with temporal vs. rate codes, there is often a tradeoff between making fewer assumptions and losing statistical power. A complete statistical treatment would consider the entire ensemble of neural responses in the decision. For example, a feature may be encoded by the relative timing of two neurons, requiring both neurons to be measured to describe the correct neural code. In practice, however, describing just one neuron's full response profile can be daunting. For this reason, population coding has proved most successful for only very well characterized areas like the retina. As was the case for rate vs. temporal codes, potential information is clearly discarded in simple averaging/pooling procedures. Neural coding is necessarily more complex than the rate-averaged single-unit level interpretation most often referenced in the literature. However, the question of how much information is carried by the relative timing of the population remains an active area of research.

Non-classical Effects: stimulus-driven and top-down

In addition to the limitations on precision from temporal and population coding, individual neurons are also subject to a wide variety of non-classical effects – effects which are not typically modeled by the standard methods. The number of deviations from the standard models can appear daunting and limitless without a unifying framework; the list includes the following: surround effects, contrast gain control, cross orientation inhibition, non-linear spatial summation, non-fourier envelope responses, spatial frequency inhibition, super-saturation, non-specific suppression, etc. Most of these effects are highly dynamic, inferential, and integrative, making them very difficult to model adequately. It is also difficult, or perhaps even misleading, to separate these effects. They can, however, be placed on a continuum between stimulus-driven and top-down effects. For more stimulus-driven effects, many fall outside the typical time frame and spatial extent of analysis. For example, stimuli normally elicit responses in V1 approximately 40-50 ms after stimulus presentation, but light-level, color, or contrast adaptation can have effects that last tens of seconds. Similarly, the area that can affect a responsive neuron is often much larger than the receptive field mapped out with simple stimuli. The number of non-classical responses in both time and space within and outside the typical window of analysis is beyond the scope of any experiment and certainly beyond the scope of this entry. The search for non-classical stimulus-driven effects is useful, and any well-formed general theory should eventually explain many of them. However, what is lacking in this area is a better theoretical explanation that makes sense of all these effects, rather than simply adding to the list with additional examples.

The effects that are more internally driven fall under many different interpretations – feedback, attention, or top-down influence. The neural code can adjust based on the

goals or collective experiences of the animal. Attention directed at specific features or locations can affect neural responses in areas as early as V1. This modulation comes from feedback in higher brain areas, including areas associated with other sensory modalities. Cues to direct attention are often implicit, but can be explicitly regulated in practice using arrows, auditory instruction, or other learned contingencies. In this way, V1 is not a purely visual area, considering that the precise neural code can be affected by non-visual cues. To process the vast amount of information efficiently, the neural code must adapt to cues outside the classical temporal and spatial extent, and top-down influence provides another ability to better encode what is most important to the animal.

### The Future of Neural Coding

Although the examples used in this entry involve the early visual areas, many of the universal techniques to interpret the neural code were discussed. Some of the tradeoffs will always be present – for example, the tradeoff between increasing statistical power with more specific stimulus assumptions (using structured artificial stimuli such as dots, bars, gratings) and losing statistical power with increased stimulus generality (noise stimuli in reverse correlation) occurs in every sensory modality and method.

Characterizations using natural scenes as stimuli appear to strike an appropriate balance by providing a wide array of ecologically relevant data that efficiently probes the space of neural responses relevant to the animal. Even simple cells in V1 are quite complex and require mathematical descriptions that are more involved than traditional models, and better characterizations are necessary for a more thorough understanding.

Computational objectives, such as compression, sparse, and independent coding offer succinct, high-level explanations of why these neurons have particular responses to stimuli. Of course, neural response characterizations and computational explanations are often idealizations. The neural code is not simply a set of isolated neurons firing at particular rates, but rather a complex combination of neurons firing at precise times in a population. Although many new methods of characterization may be quite sophisticated, they are strictly limited by the amount of data available; this limits model complexity, temporal and spatial extent of receptive fields, as well as temporal and population resolution. Also, attention and feedback modulate the neural code to selectively process the massive amount of information previously mentioned. The effects of attention make any narrow interpretation of sensory coding give way to a high-level, multimodal, and at times psychological understanding.

Making sense of the neural code is not a simple process. A wide variety of new methods and models have been introduced in recent years to address the apparent complexity of neural activity. The introduction of ideas regarding ecology and efficiency provide the opportunity to reduce this apparent complexity. They introduce a unifying framework that integrates many disparate aspects of neural coding. Neural response properties are linked to ecological goals through the study of statistical properties of the environment. This new framework allows us to address both to ‘what’ neurons respond, and more importantly ‘why’ they respond that way. In this way, physiological, computational, and ecological results all become part of a more coherent picture of methods in neural representation and coding.

Cross-References: Attention, Auditory processing: central, Neural recording, Computational approaches to perception, Cortical organization, Feature detectors: neural, Information theory, Modularity, Multimodal interactions: neural basis. Neural recording, Parallel pathways in sensory systems, Psychophysical approach to perception, Receptive fields, Reverse correlations, Visual processing, Visual scene statistics

Authors: Mark V. Albert, David J. Field

#### Further Readings:

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL & Rust

N (2005). Do we know what the early visual system does? *J Neurosci* 25, 10577–10597

Dayan, P, and Abbott, L.F. (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, MIT Press, 2001

Field DJ. (1994). What is the Goal of Sensory Coding? *Neural Computation* Vol 6: 559-601.

Rao, R, Olshausen, BA and Lewicki, (2002) *Bayesian Brain: Probabilistic Approaches to Neural Coding*, MIT Press.

Simoncelli, EP and Olshausen, BA (2001) Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193-1216.