CellPress

Probability by Time

Xaq Pitkow[1,2,*]
[1]Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA
[2]Department of Electrical and Computer Engineering, Rice University, Houston, TX 77030, USA
*Correspondence: xaq@rice.edu
http://dx.doi.org/10.1016/j.neuron.2016.10.007

In this issue of *Neuron*, Orbán et al. (2016) test whether the brain represents probabilities by sampling: do neurons interpret the world by generating causal explanations of sense data and quickly sample different interpretations over time? Orbán et al. (2016) find agreement between this model's predictions and neural data.

Neuroscience still lacks a coherent theory of neural computation. One compelling candidate is probabilistic inference (Knill and Richards, 1996), which offers a general framework for developing specific hypotheses and has been successfully used to explain a wide variety of human and animal behaviors. This provides an elegant way of understanding brain function at Marr's computational level (Marr, 2010), but we know neither the brain's algorithm nor neural mechanisms underlying such computation. In this issue of *Neuron*, Orbán et al. (2016) develop and test predictions for a neural representation of probabilities that could be used in an algorithm of probabilistic computation. This representation is known as the sampling hypothesis.

The idea that the brain computes by modeling probabilities of the world has many proponents, starting with Helmholtz (1925), who described perception as unconscious inference. In the kind of inference he advocated, "objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism" (Helmholtz, 1925). This idea is now called "analysis by synthesis" (Yuille and Kersten, 2006). The central claim is that the brain analyzes its sense data by using a mental model of the environment and then tries to find a configuration of objects in the world that could plausibly synthesize (i.e., generate or explain) that sense data. Such a mental model is useful because while the brain can never directly observe these objects, it can create an internal model (called a "generative model") that would generate its range of experiences and select good interpretations from that repertoire.

According to generative models, neural activity is supposed to emulate causal variables that explain or generate the observed sensory data. For example, an active orientation-selective neuron in primary visual cortex would signal that its preferred orientation accounts for an image patch with an edge feature. Furthermore, neurons in higher areas might signal that a particular object could explain the broad pattern of visual inputs and would interact with lower-level neurons that explain finer sensory details.

This line of reasoning about neural computation may feel backward to many who are accustomed to thinking about the brain mechanistically, for instance in terms of receptive fields and activation functions. After all, it is the stimulus which activates the neurons and not the other way around.

These two perspectives are fully compatible, however, and are just two descriptions of the same activity patterns. The generative account even entails that neurons must have receptive fields, because the neurons would be activated whenever the stimulus contains patterns that those neurons' preferred features can explain. Generative models have additional power because they can predict what types of object properties should be useful in a given natural environment.

Note that analysis by synthesis does not intrinsically require probabilistic computation, using explanations that are sensitive to uncertainty. However, even with good mental models, sense data is always uncertain. Moreover, imperfect models can increase uncertainty further (Beck et al., 2012). For this reason, the brain benefits from representing and weighing uncertainties for its hypotheses.

How might the brain do this? One convenient classification of probabilistic representations is as spatial or temporal (Savin and Denève, 2014).

In spatial representations of uncertainty, such as probabilistic population codes (Ma et al., 2006), the spatial pattern of neural activity encodes a probability distribution.

Others advocate instead for a temporal representation of probabilities. Here is where the sampling hypothesis enters (Hoyer and Hyvärinen, 2002; Orbán et al., 2016). In this model, the brain has only a point estimate of the latent variables at a single time, and uncertainty is defined by the variety of interpretations sampled across time: uncertainty about the world is represented by temporal variability in neural responses.

Sampling has been invoked to explain behavioral effects, such as bistable perception (Moreno-Bote et al., 2011) and decision biases (Vul et al., 2014). Here Orbán et al. (2016) show that sampling also accounts for many neural response properties.

In order to make concrete predictions for neural activity, Orbán et al. (2016) propose a model that relates neural activity to samples in a specific generative model for vision. They hypothesize that each neuron in primary visual cortex corresponds to one image feature (an oriented edge patch or Gabor function) that could explain the visual input and that the neuron's membrane potential corresponds to the amplitude of this feature (Figures 1A–1C). When the membrane potential hovers near one value, then this indicates that the brain is confident about the amplitude of this feature. When the membrane potential fluctuates widely, then this indicates the brain is very uncertain about

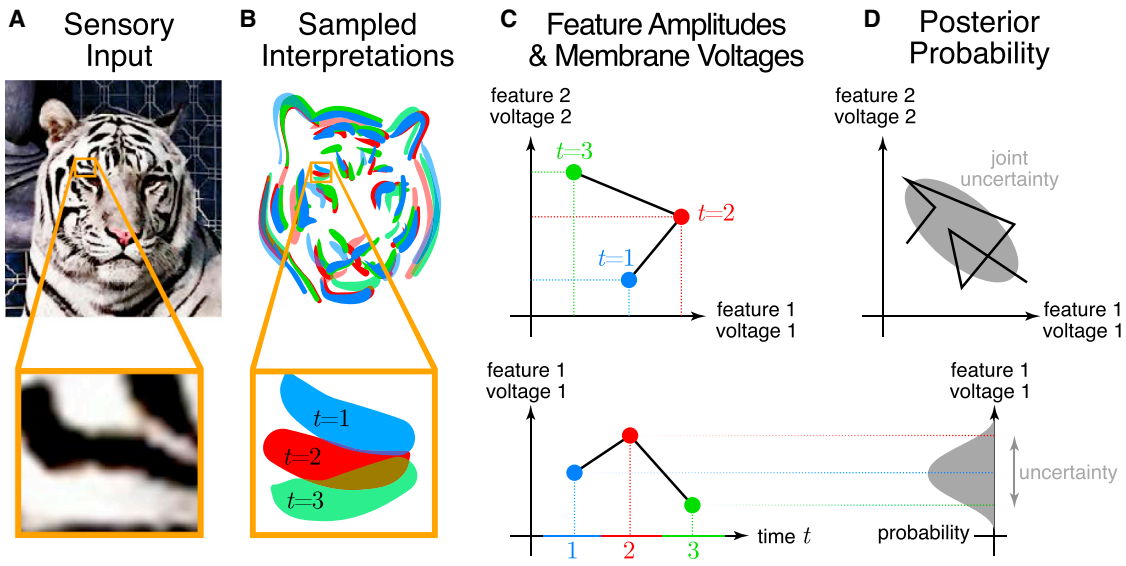

Figure 1. Cartoon of the Sampling Hypothesis

The sensory input (A) is explained or interpreted as superpositions of visual features (B). Each neuron specifies one feature, and its membrane voltage determines the amplitude of that feature. In the illustration, feature amplitude is depicted by opacity and time by color (C). Different interpretations of the image are sampled sequentially and appear as feature amplitudes—or equivalently in this sampling model, voltages—that change over time. The set of possible interpretations determines the posterior probabilities (D) both over multiple features jointly (top) and over individual features (bottom). Photo credit: Xaq Pitkow.

what amplitude best explains the visual input.

Just as variability in single membrane potentials reflects the probability distribution over amplitudes of the corresponding feature, covariability between multiple neurons reflects joint probabilities over the amplitudes of multiple features (Figures 1C and 1D).

When there is no visual stimulus, the brain exhibits spontaneous activity. Yet, we recognize darkness as dark and do not generally hallucinate complex patterns. According to the model of Orbán et al. (2016), this is because the brain's internal model includes a global contrast variable that scales all features together (Wainwright and Simoncelli, 2000). This contrast variable explains sense data in the dark, freeing the neural activity to explore random patterns without causing random percepts.

In fact, according to this sampling model, spontaneous activity patterns in the dark are samples generated from the brain's prior probability over possible world configurations. Interestingly, the spontaneous activity should match the evoked activity averaged over all natural images, a prediction that enjoys some empirical support (Berkes et al., 2011; Orbán et al., 2016; but see Okun et al., 2012).

One thing that makes predictions of the model difficult to test is that the generative causes of interest are supposed to be membrane potentials. These are much harder to measure, particularly simultaneously, than neural spiking. Orbán et al. (2016) therefore describe a simple nonlinear model to relate voltage to spiking activity.

By combining their sampling model with this spiking model, Orbán et al. (2016) can then directly compare several properties of their sampling representation to neural properties extracted from publicly available experimental data. In each case, they process their simulated data in the identical ways as the published neurophysiological data.

These comparisons show that their model naturally reproduces the observed Fano factors, which are greater for spontaneous activity than for evoked activity. The model demonstrates the same match between signal correlations, noise correlations, and spontaneous correlations that is seen in neural data. Again, like the neural data, their model also exhibits sparser and more reliable responses when the visual stimuli provide greater context for a given image patch. The agreements are remarkable, especially given that there are just a handful of free model parameters that are fixed across all comparisons against diverse datasets. These successes of the sampling hypothesis should certainly encourage further investigations.

How could the brain generate samples? The work of Orbán et al. (2016) is agnostic about the neural mechanisms. But the basic notion is that the uncertainty associated with each possible input is encoded implicitly in the network connectivity and stochastic neural response properties. Samples of the target probability distribution are then created by the circuit dynamics.

Note that any representation could be trivially considered probabilistic just by applying Bayes' rule. But the value of a probabilistic representation depends on whether subsequent computations take advantage of that representation. Thus, it is crucial to ask, how would the brain use these samples?

One natural use of samples is integration over a probability distribution. Samples could be passed through some nonlinearity representing a reward function. Decision-making circuits could then integrate the results over time to obtain the expected reward and identify the best action as the one that maximizes the expected reward.

In contrast, temporal integration of evidence does not seem to be a natural use of sampling, at least in its current incarnation. This is because percepts become more reliable at greater viewing durations, whereas sampling longer does not reduce uncertainty, but instead just specifies it better. Neural responses from a sampling process could be simply integrated over time, which would improve the reliability of derived estimates, but then this would be a departure from the reported sampling representation. It is appealing and certainly worthwhile to search for a general-purpose neural architecture. However, since different tasks present different computational demands, it remains possible that the brain uses different representations for different tasks.

Spatial and temporal representations of probabilities each have experimental support in different aspects, but each faces major difficulties in representing arbitrary multivariate distributions: spatial codes require a number of neurons that grow exponentially with the number of variables, whereas temporal codes can require arbitrarily long times to sample all probable states. It may be that some hybrid spatiotemporal representation (Lee and Mumford, 2003; Savin and Denève, 2014) can capture the best aspects of each model.

Of course, the natural world is not arbitrary, but highly structured, a fact that any algorithm must exploit. As a simple foray for considering the effects of a hierarchically structured environment, Orbán et al. (2016) vary an image aperture to provide more image context. They show that the consequences are the same as for an increase in contrast, both in their model and in neural data. Simple cases like this are crucial stepping stones for testing the sampling hypothesis. Ultimately, any theory will have to grapple with how the brain handles more complex, statistical structure as well (Haefner et al., 2016).

Sampling is a common and broadly applicable class of techniques in machine learning for computing with complex probabilistic models. Yet, accurate sampling remains a major practical challenge, and new variations are developed continually. If the brain does compute with sampling, and has found a clever trick for doing so efficiently, then we could adapt that trick to our own technology. Perhaps sampling can go some way toward explaining how we think and could even help us think better.

REFERENCES

Beck, J.M., Ma, W.J., Pitkow, X., Latham, P.E., and Pouget, A. (2012). Neuron *74*, 30–39.

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Science *331*, 83–87.

Haefner, R.M., Berkes, P., and Fiser, J. (2016). Neuron *90*, 649–660.

Helmholtz, H. (1925). Optical Society of America *3*, 318.

Hoyer, P.O., and Hyvärinen, A. (2002). S. Becker, S. Thrun, and K. Obermayer, eds. Advances in Neural Information Processing Systems (NIPS 2002) 15, 293–300.

Knill, D.C., and Richards, W. (1996). Perception as Bayesian Inference (Cambridge University Press).

Lee, T.S., and Mumford, D. (2003). J. Opt. Soc. Am. A Opt. Image Sci. Vis. *20*, 1434–1448.

Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Nat. Neurosci. *9*, 1432–1438.

Marr, D. (2010). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (MIT Press).

Moreno-Bote, R., Knill, D.C., and Pouget, A. (2011). Proc. Natl. Acad. Sci. USA *108*, 12491–12496.

Okun, M., Yger, P., Marguet, S.L., Gerard-Mercier, F., Benucci, A., Katzner, S., Busse, L., Carandini, M., and Harris, K.D. (2012). J. Neurosci. *32*, 17108–17119.

Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. (2016). Neuron *92*, this issue, 530–543.

Savin, C., and Denève, S. (2014). Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, eds. Advances in Neural Information Processing Systems (NIPS 2014) 27, 2024–2032.

Vul, E., Goodman, N., Griffiths, T.L., and Tenenbaum, J.B. (2014). Cogn. Sci. *38*, 599–637.

Wainwright, M.J., and Simoncelli, E.P. (2000). Scale mixtures of Gaussians and the statistics of natural images. S.A. Solla, T.K. Leen, and K. Müller, eds. Advances in Neural Information Processing Systems (NIPS 1999) 12, 855–861.

Yuille, A., and Kersten, D. (2006). Trends Cogn. Sci. *10*, 301–308.