Seed Magazineabout

SEEDMAGAZINE.COM *July 20, 2009*

The Prophetic Brain

Universe in 2009 by Karl Friston / January 27, 2009

The **commonly held belief** that information from the outside world impinges upon our brains through our senses to cause perception and then action **now appears to be false.**

Over the past decade, neuroscience has revealed that rather than acting as a filter that simply maps sensation onto action, the brain behaves like an “inference machine” that tries to discover patterns within data by refining a model of how those patterns are likely to be generated. For instance, depending on whether the context is a crowded concert hall or a deserted forest, a sound can be perceived as either a human voice or the wind whistling through trees. The pioneering German physicist Hermann von Helmholtz articulated this idea as early as 1860, when he wrote of visual perception that “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.” Now a unified understanding of how the brain makes and optimizes its inferences about the outside world is emerging from even earlier work—that of the 18th-century mathematician Thomas Bayes.

Bayes developed a statistical method to evaluate the probability of any given hypothesis being true under changing conditions. The concept is straightforward: The probability of two things happening together is the probability of the first given the second, times the probability of the second. This allows the certainty of a single inference to be weighed according to how much additional evidence exists at any particular time. The “Bayesian” approach has emerged in many guises over the past century and has proved very useful in computer science applications like machine learning.

Since at least the 1980s, neuroscientists have speculated that the brain may use Bayesian inference to make predictions about the outside world. In this view, the brain estimates the most likely cause of an observation (that is, sensory input) by computing the probability that a particular series of events generated what was observed—not unlike a scientist who constructs a model to fit his or her data. This probability is a mathematical quantity we call the “evidence.” But evaluating the evidence for most realistic models requires calculations so intricate and lengthy they become impractical. This would be particularly problematic for the brain, which must constantly make split-second decisions. Fortunately, there is an easier way. In 1972 the American physicist Richard Feynman devised an elegant shortcut to calculate the evidence using something called a “free-energy bound.” Freeenergy is a concept from statistical thermodynamics— it is essentially the energy that can be used for work within a system once that system’s entropy, or useless energy, has been subtracted.

Feynman’s basic idea was simple: Instead of trying to compute the evidence explicitly, just start with a quantitative guess about the causes, which we will call a “representation,” and then adjust the representation until it minimizes the free-energy of the data. Feynman exploited the fact that the freeenergy is, by construction, always greater than the negative logarithm of the evidence, a mathematical quantity we will call “surprise.” In other words, the free-energy is an *upper boundary* upon surprise (remember this—we’ll come back to it later). So by changing the representation to *minimize* freeenergy, the representation becomes the most likely cause of whatever sensory inputs make up an observation, and the free-energy becomes the evidence itself. The machine-learning community has used this approach with great success, leading many researchers to wonder: If minimizing free-energy is so effective in allowing statistical machines to perceive and “learn” about their surroundings, could the brain be taking similar shortcuts?

In this formulation, a “representation” is simply a quantitative guess about the likely cause of a sensory observation. To understand representation in the brain, imagine you are in a bar having a conversation. The sounds you hear have no meaning beyond being the product of someone speaking. Your brain must first represent the deeper cause of the sounds (in this case, the concepts and words that make up the speech) via its internal variables like the activity of neurons and the strengths of connections between them. Only then can you infer any meaning. What would this process look like?

The emerging picture is that the brain makes its inferences by minimizing the free-energy of messages passing between hierarchical brain regions. Imagine the brain as an onion, where meaningful exchanges with the outside world take place on its surface (the outer sensory layer). Information from these exchanges passes on to “higher” levels (those responsible for cognitive functions) through “bottom up” connections. The higher levels respond with “top down” messages to the lower levels. This reciprocal exchange repeats itself hierarchically, back and forth, layer by layer, until the highest level (at the center of the onion, or front of the brain) becomes engaged. Only then will you consciously register a perception. In this scheme, the free-energy is essentially the collective prediction error over all levels of the hierarchy: Top-down cognitive messages provide predictions based on representations from above, and lower sensory levels reciprocate with bottom-up prediction errors. These “error messages” drive encoded representations (such as neuronal activity) to improve the predictions for lower levels (that is, to reduce free-energy).

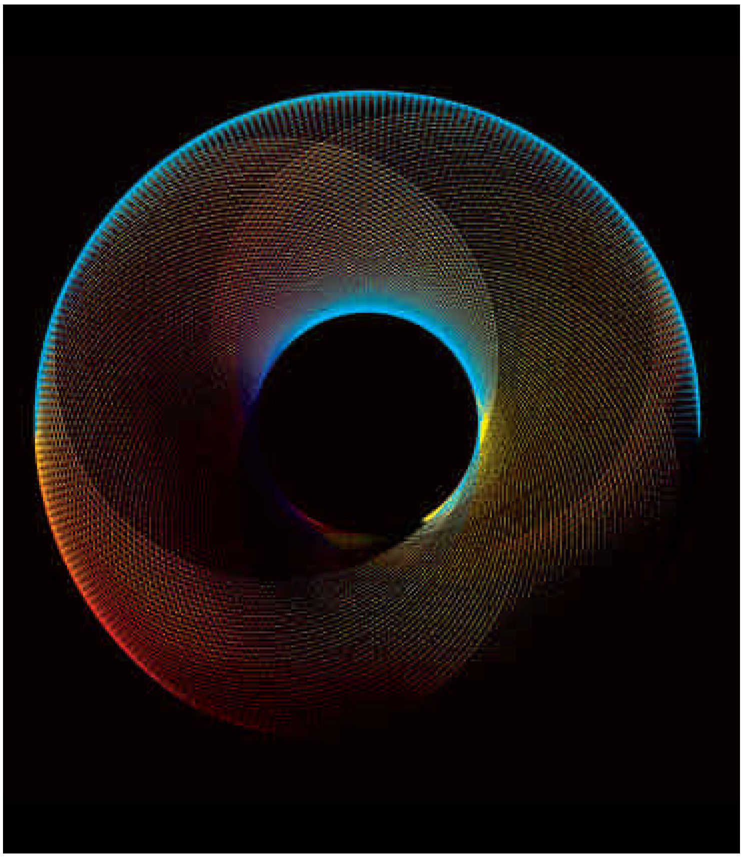

Illustration by Andy Gilmore

For example, in your hypothetical bar conversation, no matter how ambiguous the acoustics, you are more likely to hear “credit crunch” as opposed to “credit brunch.” Here the high-level conceptual representation “credit crunch” provides contextual constraints on the words, which restrict the sounds predicted or heard, namely “c,” not “b.” If the bar is very noisy, you may find yourself watching your friend’s mouth closely. This is because the cause (speaking) allows you to make both acoustic and visual predictions. Hierarchical optimization allows sounds to help you see and sights to guide hearing—binding different sensations into a coherent perceptual framework. This recurrent messagepassing leads to the self-organized brain dynamics that support perception and recognition.

The perspective afforded by this hierarchical Bayesian formulation is especially important for neuroscience, because hierarchy *is* a key architectural principle of brain anatomy—our brains *are* organized in successive layers, and we *can* measure the neural activity encoding prediction errors and representations. But this is not the end of the story. What follows is a new theory that considers what would happen if the free-energy principle applied not just to *perception* but to *action* as well.

Let’s begin with the notion of an ensemble density—a probability distribution of the states you or I can occupy. Imagine I had 100 million copies of you, at different times in your daily life. If I could measure all your sensory states, I could construct a sample density or histogram that reflected the probability of your being in any particular state. Critically, for you to exist, the number of states you occupy must be small in relation to all possible states. For example, your temperature will always be in a certain range. Mathematically, this means your ensemble density has low entropy. Here, we meet a characteristic of adaptive biological agents (like you and I) in that they seem to resist the second law of thermodynamics (a universal tendency to disorder) by minimizing the entropy of their ensemble densities. What does minimizing entropy mean? It simply means that you will, on average, avoid surprising or improbable states (i.e., you will not find yourself at the bottom of the ocean or suddenly engulfed in flames). Though arcane, this implies something quite fundamental: To exist, you must avoid surprising states.

ON THE BLOGS

Last year, ScienceBlogger and cognitive neuroscientist Chris Chatham delved into aspects of Karl Friston's own scientific research, including:

How Karl Friston's theoretical models of brain function relate to object recognition and visual processing.

How to determine in an fMRI experiment whether two different tasks are activating the same region of the brain.

Adaptive agents like us are open systems that exchange with their environment. The environment acts on us, which produces sensory impressions, and we act on the environment to change its states: If you see an apple on a table, you can reach out to pick the apple up. If we can change the environment that causes sensory input, then, in principle, we can act to suppress surprising input. But there is a problem: How do we compute surprise? In fact, we do not need to compute surprise at all. Returning to Feynman's elegant methodology, all we need to do is to minimize free-energy, because free-energy is an upper boundary on surprise. This means that free-energy can be used not only to optimize perception, but also to prescribe action. This is the basis of the freeenergy principle, which states that *all quantities associated with an agent will change to minimize free-energy*. This line of reasoning prescribes an intimate relationship between perception and action, where both work in concert to suppress free-energy (that is, to minimize prediction errors or surprise) in our sensory experiences. In other words, we will actively sample sensory data so that it conforms to our expectations; we will constantly alter our relationship with our environment so that our expectations become self-fulfilling prophecies. A simple example of this is turning one's head to get a better view of what seems to be a familiar face in peripheral vision, but this principle may encompass our entire navigation of the world to avoid the unexpected.

In terms of neuroscience, the key issue is not so much the information theoretic principles above, but how the brain *realizes* them. Multiple predictions follow from these ideas. For example, brain systems should be deployed hierarchically and connected reciprocally. Forward connections should be largely linear in their influences, whereas backward connections should embody the nonlinearities inherent in the causal structure of the world. We would expect that predictable stimuli evoke smaller responses, and unexpected stimuli larger ones. Scientists are now starting to confirm these conjectures with brain mapping, by comparing brain responses with stimuli that are coherent or incoherent, predictable or unpredictable. This principle also has implications beyond neuroscience, in the sense that it applies to all biological agents. Could single-cell organisms use the concentration of metabolites and kinetic rate-constants (as opposed to neuronal activity and connection strengths) to encode their implicit representations? In this speculative case as well as with the brain, the great challenge is to find the mapping between the internal states of a phenotype and representations that this theory mandates.

Returning to statistical machines, from which much of this work emerged, the theory suggests a profound revision of current approaches to reinforcement learning and optimal control in engineering artificial neural networks. It should be possible to teach automata (such as robots) complex adaptive behaviors by simply exposing them to a controlled environment (like a classroom), then returning them to their normal surroundings to seek out the new states they have learned to expect. The limitations of this approach are difficult to predict, but further synergy between theoretical neurobiology and machine learning, between a deeper understanding of our own minds and those we wish to create, appears inevitable. *—Karl Friston is the scientific director of the Wellcome Trust Center for Neuroimaging.*