In: SAGE Encyclopedia of Perception (2009). *Goldstein EB ed., 201-5.*

BAYESIAN APPROACH
Wei Ji Ma, Baylor College of Medicine

The Bayesian approach to perception focuses on whether the computations that the brain performs in perceiving the world can be described as forms of *Bayesian inference*. Bayesian inference is the process of optimally extracting information from noisy inputs and, if needed, combining this information with prior knowledge. The Bayesian approach to perception typically involves mathematically precise modeling of perceptual experiments in humans. Central to Bayesian inference is *Bayes' rule* (also called *Bayes' theorem*), named after the 18^{th}-century British mathematician and Presbyterian minister Thomas Bayes. Bayes' rule is a very general equation directly derived from the basic tenets of probability theory, but in the context of perception, it is based on the fact that the brain's observations of physical stimuli (such as the orientation of a line or the identity of an object) are noisy. It states that the probability that a stimulus had a particular value given a set of noisy observations (the *posterior probability* of the stimulus) is proportional to the probability that a stimulus of that value generated those observations (the *likelihood* of the stimulus), multiplied by the probability of the stimulus value independent of the observations (the *prior probability* of the stimulus). In mathematical notation,

$$p(\text{stimulus} \mid \text{observations}) \propto p(\text{observations} \mid \text{stimulus}) \times p(\text{stimulus}) \quad (1)$$

In this equation, p stands for probability, "|" means "given", and "α" means "proportional to". (In more abstract notation, Bayes' rule is written as $p(A|B) \propto p(B|A)\, p(A)$, where A and B are random variables that could refer to the stimulus and the observations, respectively.) We will discuss several concrete examples in this article. Ernst Mach (1886) and Hermann von Helmholtz (1891) were among the first to apply the idea of Bayesian inference to sensory perception, although a very similar notion can be found in the writings of Pierre-Simon Laplace (1825). Strong experimental evidence for Bayesian inference in perception started emerging only in the 1980s, and we will discuss some of it here. The Bayesian approach to perception should not be confused with Bayesian data analysis, which is an area of statistics with applications in neuroscience.

Perception as information extraction

Our percept of the world is a complex composite of pieces of uncertain knowledge extracted from noisy and ambiguous sensory inputs. The brain is very different from a recording device such as a camera, which accurately captures the light in a visual scene but will not be able to tell, for instance, whether there is an animal in the scene. Indeed, in computer vision, sophisticated algorithms are used to extract that information. To humans and other animals, such higher-level knowledge can be of great ecological importance – for example, an animal in the scene could be a predator. To extract information from sensory inputs, the brain has to perform many different manipulations, including combining signals that come from the same source, making decisions, planning movements, allocating attention, and storing and retrieving information. Manipulations of sensory inputs that aim to extract behaviorally relevant information can be called *perceptual computations*. It is a key question in neuroscience and cognitive science what algorithms and mechanisms the brain has developed to perform perceptual computations in the presence of uncertainty.

Uncertainty in perception

Sensory information is uncertain. A lion chasing a giraffe must judge its position, speed, and direction of motion to decide when and how to attack, but because the giraffe is running fast, these judgments come with large uncertainty. A basketball player trying to pass the ball must keep track of the positions of his teammates and of the members of the opposing team, sometimes only seeing them out of the corner of his eye. When someone approaching you looks familiar, you need to use visual information to decide whether to greet your friend or avoid an awkward moment. When you are hiking in the forest and need to decide whether to jump over a stream, you have to judge the width of the stream as well as how far you can jump. When you are driving on the freeway and the car ahead of you brakes, you determine how to react based on uncertain variables such as your car's distance to it. The central idea of the Bayesian approach is to acknowledge this omnipresent uncertainty and ask how the brain can take it into account optimally when performing perceptual computations.

An analogy: the apple tree

Because sensory information is noisy, the brain constantly has to make guesses about the state of the world. Clearly, some guesses are better than others. Computing the goodness of possible guesses is known as Bayesian inference. Imagine you are shown the locations of some apples lying scattered in a field. You are told that the apples fell from a tree recently and are asked to guess the location of the tree. Your best guess will probably be somewhere in the middle between the apples, since you know that apples don't fall far from the tree and tend to fall roughly evenly in all directions. Not only will you be able to report a best guess of the tree location, you will also have an idea how probable each other location in the field is. For example, the farther outward a location is, the less you will believe that the apple tree was there. This changes if you know that there was a strong wind blowing from the West when the apples fell. Your best guess about the location of the tree would then be more to the East than it was before, and probabilities assigned to other locations will change accordingly. If you have additional information, for example that some areas of the field are swampy, you will be able to adjust your probabilities, because you know apple trees tend to grow on solid ground.

This example illustrates that one can use noisy observations (the locations of the apples) to obtain the probability of each of their possible causes (apple tree positions). However, for this to work, you need to know how a given cause generates observations (for example, that apples don't fall far from the tree). You might also have prior information about causes (such as that apple trees tend to grow on solid ground). Using these pieces of knowledge when you are given a particular set of observations, will allow you to infer the goodness of possible guesses about the stimulus. This is exactly the logic expressed by Bayes' rule – in this example, it would state that the probability of an apple tree location given locations of apples in the field can be computed by multiplying the probability of those apple locations given that tree location with the prior probability of that tree location. This multiplication is followed by a *normalization*, which is the division by a constant factor to ensure that the total probability of all possible causes is equal to 1 – a defining property of a probability distribution. One can write this as *p*(tree location | apple locations) α *p*(apple locations | tree location) x *p*(tree location).

Obviously, Bayes' rule is not just about apple trees. It is very general and widely used in fields as different as statistics, economics, computer science, computational biology, and nowadays also neuroscience. Moreover, it has practical applications, such as in Bayesian spam filters, which calculate the probability that an email is spam based on the words it contains. They do this by multiplying the probability of finding those words in a spam email by the prior probability that any email is spam, followed by a normalization: *p*(spam | words) α *p*(words | spam) x *p*(spam).

Bayesian inference in perception

We will now discuss what Bayes' rule means in the context of perception. The apple tree in the example above is analogous to a feature of an event or object in the world, often called the *stimulus* (in a somewhat confusing terminology, the event or object itself is sometimes also called the stimulus; it is then understood which of its features is of relevance). This can be a simple variable like the orientation of a line, the color of a surface, or the direction of motion of an object. It can also be more complicated, like whether a desired object is present in a scene, what your coworker just said, the identity of the person approaching you, or how safe it is to change lanes on the freeway.

The apple locations are analogous to observations of the stimulus made by the brain. There are different ways of conceptualizing these. In their simplest form, the observations are variable measurements of the stimulus. For example, when the stimulus is a line oriented at 60°, a measurement of its orientation on a single trial could be 58°, 61°, etc. This is sometimes called the *internal representation* of the stimulus. It is often assumed that internal representation of a stimulus is distributed around the true value of the stimulus, according to a roughly Gaussian (normal) distribution. Alternatively, the observations could be the variable activities of neurons in a population responding to the stimulus. For example, if the stimulus is an oriented line, then the observations could be the numbers of spikes emitted by the neurons in primary visual cortex (V1) that are sensitive to orientation at that location. These numbers are variable, because not every time you present the same stimulus, you get the same pattern of response. Which description of the observations is most appropriate depends mostly on whether you study a perceptual phenomenon from a behavioral or a neural point of view: psychologists tend to use noisy measurements, while neuroscientists tend to use neural activities. As of yet, most Bayesian models of perception are at the behavioral level.

Now that we have described what the stimulus and the observations are, we can formulate the central hypothesis of the Bayesian approach in perception: in many perceptual tasks, the brain computes the posterior probability distribution over the relevant stimulus variable based on the observations it has available, as expressed in Equation (1). One can think of this equation as a formalized description of a key task that the brain faces: to infer aspects of the world based on noisy sensory input. The lion infers in which direction the giraffe is moving, the basketball player where his teammates are, and the forest hiker whether he can jump over the stream. From the point of view of the observations, *p*(observations | stimulus) is sometimes called the *noise model*, the *variability* in the observations, or their *generative model*. In the apple tree analogy, the clue that there is a strong wind from the West is an element of the generative model. In the example of the oriented line, the Gaussian distribution or the form of the variability of the population of V1 neurons constitutes the generative model. When the observations are the activities of a population of neurons, one can think of that population as *encoding* the probability distribution over the stimulus (see Figure 1).

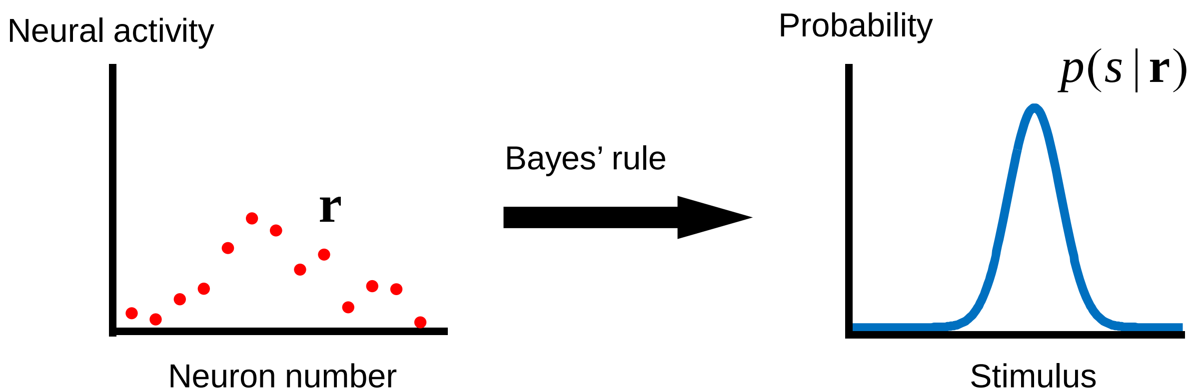

Figure 1: Schematic illustration of Bayes' rule in perception (neural version). A stimulus (e.g. orientation of a line) elicits activity in a population of neurons (left), whose activities (e.g. spike counts) we collectively denote by **r**. Based on these noisy observations, Bayes' rule computes the probability distribution $p(s|\mathbf{r})$ over the stimulus (right), providing not only the most likely value of the stimulus, but also its uncertainty. One can think of the probability distribution being *encoded* in the pattern of activity.

The posterior probability distribution, *p*(stimulus | observations), reflects the uncertainty about the stimulus: if this distribution has a single sharp peak, it means that based on the observations, only one of very few stimuli are plausible, and uncertainty is small. For example, when you are driving on the freeway on a sunny day and there is another car a short distance in front of you, the posterior probability distribution over your distance to it will be narrow. If, on the contrary, the distribution has a flat and broad peak, or multiple peaks of similar height, it means that based on the observations, many stimuli are plausible, and uncertainty is high. For example, if the same car is far ahead of you at nighttime, the posterior probability distribution will be broad. This is an important point, because it means that, through Bayes' rule, the brain can encode not only the most likely value of a stimulus (compare: the most likely location of the apple tree, the most likely distance to the other car), but also the uncertainty about that value (compare: how large the area is where you believe the apple tree could still stand, how wide a range of distances to the other car is plausible).

Bayesian inference in perception is not always, and even seldom, a matter of perceiving a single stimulus. As we discussed in the section "Perception as information extraction", perceptual computation can involve elaborate manipulation of sensory information, which in many cases has to be done in multiple stages. For example, when you are deciding whether there is a predator hiding the bushes, your brain has to infer not only the features of objects in the scene, but also make the higher-level judgment about animal presence. In such situations, Bayesian inference can become very complex, but the key feature remains that at every stage, a probability distribution is computed over the variables at that stage. Using Bayes' rule to compute with uncertainty and probability during a perceptual task guarantees the best possible performance; therefore, Bayesian inference is often referred to as *optimal inference,* or "optimally taking into account uncertainty". The Bayes-optimal strategy ensures that an organism makes fewer mistakes and is more efficient in perceiving the world and generating behavior.

Experimental evidence

Perception researchers have found a large amount of evidence that the brain performs Bayesian inference, i.e., computes posterior probability distributions over stimuli. This evidence is commonly provided by showing that those probabilities are used by humans in solving certain behavioral tasks. No experimental paradigm has contributed more to this than that of cue combination. The basic idea of this perceptual computation is that an observer is presented with two or more cues providing information about the same stimulus, and has to estimate the stimulus. For example, you might have a cat who likes to hide in the backyard. To find her, you use the movement you see as well as the faint meowing you hear. Because of noise, both in the world and inside the brain, cues from the same source are often not in complete agreement. For example, the best guess about the location of the cat based on its meowing may be slightly different from the best guess based on its image. One way to combine such cues would be to take the average of these guesses. However, this is clearly not a good idea when one cue is much less certain than the other: if the cat is barely visible, the best combined guess may be very close to the one based on what you hear. Indeed, the formalism of Bayesian inference applied to cue combination predicts that the best combined estimate is a specific combination of the individual guesses that gives less weight to the less certain cue. This prediction has been tested in a wide variety of cue combination experiments in which the disagreement between two cues is artifically varied. The general conclusion has been that humans do take into account uncertainty in a close to Bayes-optimal way when combining cues. Evidence for Bayesian inference has been found in many other areas of perception as well, including decision making, visual motion estimation, color perception, and sensorimotor behavior.

Neural basis
Even though Bayesian inference has become a fixture in behavioral descriptions of perception, the neural basis of Bayes-optimal computations has only recently started to receive scrutiny. The challenge here is to discover which operations the brain performs on neural activity so as to produce behavior that is close to Bayes-optimal. This problem is directly related to one of the central questions in neuroscience, namely that of the relation between brain and behavior. Eventually, predictions from this approach might be tested in physiology. Research into Bayesian inference in perception is expected to be of wide interest in the coming years.

Wei Ji Ma

Related topics
Theoretical approaches
Computational approach
Multimodal interactions
Neural representation/coding
Statistical learning
Signal detection theory and procedures

Further readings

Doya, K., Ishii, S., Pouget, A., & Rao, R.P.N. (2007). *Bayesian brain: probabilistic approaches to neural coding.* Cambridge, MA: MIT Press.
Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429-433.
Knill, D.C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences* 27, 712-719.
Kersten, D., Mamassian, P., & Yuille, A. (2005). Object perception as Bayesian inference. *Annual Review of Psychology* 55, 271-304.
Kording, K. (2007). Decision theory: what "should" the nervous system do? *Science* 318: 606-610.
Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience* 9, 1432-1438.
Rao, R.P.N., Olshausen, B.A., & Lewicki, M.S. (2002). *Probabilistic models of the brain: perception and neural function*. Cambridge, MA: MIT Press.