The Bayesian brain: the role of uncertainty in neural coding and computation

David C. Knill and Alexandre Pouget

Center for Visual Science and the Department of Brain and Cognitive Science, University of Rochester, NY 14627, USA

To use sensory information efficiently to make judgments and guide action in the world, the brain must represent and use information about uncertainty in its computations for perception and action. Bayesian methods have proven successful in building computational theories for perception and sensorimotor control, and psychophysics is providing a growing body of evidence that human perceptual computations are 'Bayes' optimal'. This leads to the 'Bayesian coding hypothesis': that the brain represents sensory information probabilistically, in the form of probability distributions. Several computational schemes have recently been proposed for how this might be achieved in populations of neurons. Neurophysiological data on the hypothesis, however, is almost non-existent. A major challenge for neuroscientists is to test these ideas experimentally, and so determine whether and how neurons code information about sensory uncertainty.

Humans and other animals operate in a world of sensory uncertainty. Although introspection tells us that perception is deterministic and certain, many factors contribute to limiting the reliability of sensory information about the world – the mapping of 3D objects into a 2D image, neural noise introduced in early stages of sensory coding, and structural constraints on neural representations and computations (e.g. the density of receptors in the retina). Our brains must effectively deal with the resulting uncertainty to generate perceptual representations of the world and to guide our actions. This leads naturally to the idea that perception is a process of unconscious, probabilistic inference [1,2]. Aided by developments in statistics and artificial intelligence, researchers have begun to apply the concepts of probability theory rigorously to problems in biological perception and action [3–20]. One striking observation from this work is the myriad ways in which human observers behave as optimal Bayesian observers. This observation, along with the behavioral and computational work on which it is based, has fundamental implications for neuroscience, particularly in how we conceive of neural computations and the nature of neural representations of perceptual and motor variables.

Corresponding author: David C. Knill (knill@cvs.rochester.edu).

Bayesian inference and the Bayesian coding hypothesis

The fundamental concept behind the Bayesian approach to perceptual computations is that the information provided by a set of sensory data about the world is represented by a conditional probability density function over the set of unknown variables – the posterior density function. A Bayesian perceptual system, therefore, would represent the perceived depth of an object, for example, not as a single number Z but as a conditional probability density function $p(Z/I)$, where I is the available image information (e.g. stereo disparities). Loosely speaking, $p(Z/I)$ would specify the relative probability that the object is at different depths Z, given the available sensory information.

More generally, the component computations that underlay Bayesian inferences [that give rise to $p(Z/I)$] are ideally performed on representations of conditional probability density functions rather than on unitary estimates of parameter values. Loosely speaking, a Bayes' optimal system maintains, at each stage of local computation, a representation of all possible values of the parameters being computed along with associated probabilities. This allows the system to integrate information efficiently over space and time, to integrate information from different sensory cues and sensory modalities, and to propagate information from one stage of processing to another without committing too early to particular interpretations. Bayesian statisticians refer to the idea of representing and propagating information in the form of conditional density functions as belief propagation, and this approach has been highly successful in designing effective artificial vision systems [21–23].

To illustrate the basic structure of Bayesian computations, consider the problem of integrating multiple sensory cues about some property of a scene. Figure 1 illustrates the Bayesian formulation of one such problem – estimating the position of an object X from visual and auditory cues V and A. The goal of an optimal, Bayesian observer would be to compute the conditional density function $p(X/V,A)$. Using Bayes' rule, this is given by

$$P(X/V,A) = p(V,A/X)p(X)/p(V,A) \qquad \text{(Equation 1)}$$

where $p(V,A/X)$ specifies the relative likelihood of sensing the given data for different values of X and $p(X)$ is the prior probability of different values of X. Because the noise
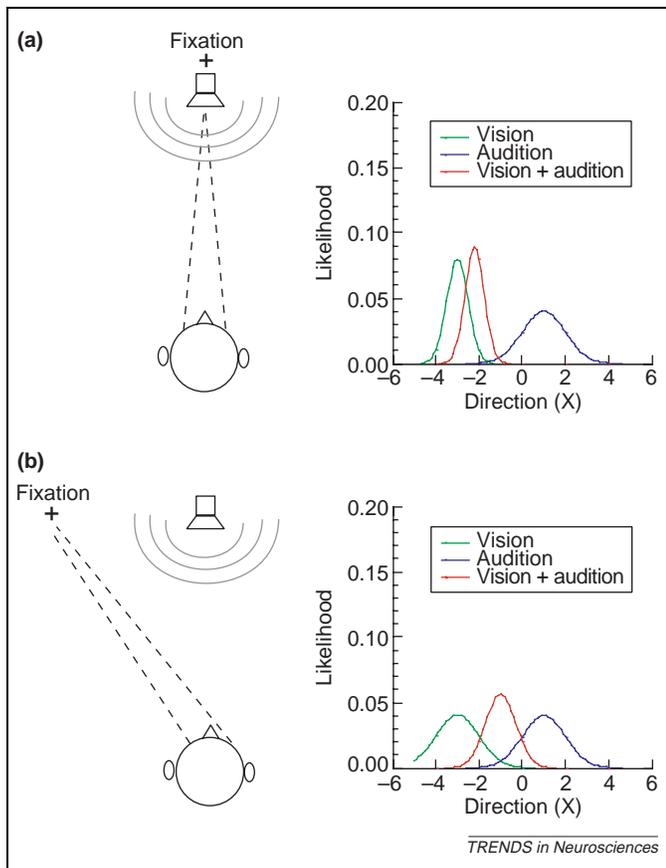
Figure 1. Two examples in which auditory and visual cues provide 'conflicting' information about the direction of a target. The conflict is apparent in the difference in means of the likelihood functions associated with each cue, although the functions overlap. Such conflicts are always present, owing to noise in the sensory systems. To integrate visual and auditory information optimally, a multimodal area must take into account the uncertainty associated with each cue. **(a)** When the vision cue is most reliable, the peak of the posterior distribution is shifted toward the direction suggested by the vision cue. **(b)** When the reliabilities of the cues are more similar, for example when the stimulus is in the far periphery, the peak is shifted toward the direction suggested by the auditory cue. When both likelihood functions are Gaussian, the most likely direction of the target is given by a weighted sum of the most likely directions (μ) given the visual (V) and auditory (A) cues individually: $\mu_{V,A} = w_V\mu_V + w_A\mu_A$. The weights (w) are inversely proportional to the variances of the likelihood functions.

sources in auditory and visual mechanisms are statistically independent, we can decompose the likelihood function into the product of likelihood functions associated with the visual and auditory cues, respectively:

$$P(V,A/X) = p(V/X)p(A/X) \qquad \text{(Equation 2)}$$

$p(V/X)$ and $p(A/X)$ fully represent the information provided by the visual and auditory data about the position of the target. The posterior density function is therefore proportional to the product of three functions: the likelihood functions associated with each cue and the prior density function representing the relative probability of the target being at any given position. An optimal estimator could pick the peak of the posterior density function, the mean of the function or any of several other choices, depending on the cost associated with making different types of errors [24].

For our purposes, the point of the example is that an optimal integrator must take into account the relative uncertainty of each cue when deriving an integrated estimate. When one cue is less certain than another, the integrated estimate should be biased toward the more

reliable cue. Assuming that a system can accurately compute and represent likelihood functions, the calculation embodied in equations 1 and 2 implicitly enforces this behavior (Figure 1). Although other estimation schemes can show the same performance as an optimal Bayesian observer (e.g. a weighted sum of estimates independently derived from each cue), computing with likelihood functions provides the most direct means available to account 'automatically' for the large range of differences in cue uncertainty that an observer is likely to face.

This is the basic premise on which Bayesian theories of cortical processing will succeed or fail – that the brain represents information probabilistically, by coding and computing with probability density functions or approximations to probability density functions. We will refer to this as the 'Bayesian coding hypothesis'. The opposing view is that neural representations are deterministic and discrete, which might be intuitive but also misleading. This intuition might be due to the apparent 'oneness' of our perceptual world and the need to 'collapse' perceptual representations into discrete actions, such as decisions or motor behaviors. The principle data on the Bayesian coding hypothesis are behavioral results showing the many different ways in which humans perform as Bayesian observers.

Are human observers Bayes' optimal?
What does it mean to say that an observer is 'Bayes' optimal'? Humans are clearly not optimal in the sense that they achieve the level of performance afforded by the uncertainty in the physical stimulus. Absolute efficiencies (a measure of performance relative to a Bayes' optimal observer) for performing high-level perceptual tasks are generally low and vary widely across tasks. In some cases, this inefficiency is entirely due to uncertainty in the coding of sensory primitives that serve as inputs to perceptual computations [6]; in others, it is due to a combination of sensory, perceptual and cognitive factors [25]. The real test of the Bayesian coding hypothesis is in whether the neural computations that result in perceptual judgments or motor behavior take into account the uncertainty in the information available at each stage of processing. Psychophysical work in several areas suggests that this is the case.

Cue integration
Perhaps the most persuasive evidence for the Bayesian coding hypothesis comes from work on sensory cue integration. When the uncertainty associated with each of a set of cues is approximated by a Gaussian likelihood function, the average estimate derived from an optimal Bayesian integrator is a weighted average of the average estimates that would be derived from each cue alone (Figure 1). The reliability of different cues changes as a function of many scene and viewing parameters (e.g. the reliability of stereo disparity decreases with viewing distance). When these parameters vary from trial to trial in a psychophysical experiment, an optimal Bayesian observer would appear to weight cues differently on different trials. Studies of human cue integration, both within modality (e.g. stereo and texture) [26–28] and

across modality (e.g. sight and touch or sight and sound) [29–32], consistently find cue weights that vary in the manner predicted by Bayesian theory. Although these results could be accounted for by a deterministic system that adjusts cue weights as a function of viewing parameters and stimulus properties that co-vary with cue uncertainty, representing and computing with probability distributions (Figure 1) is considerably more flexible and can accommodate novel stimulus changes that alter cue uncertainty.

Non-linear Bayesian estimation

One of the strongest computational arguments for representing density functions in intermediate perceptual computations is that they are often not Gaussian, so that simple linear mechanisms do not suffice to support optimal Bayesian calculations. Non-Gaussian likelihood functions arise even when the sensory noise is Gaussian as a result of the nonlinear mapping from sensory feature space to the parameter space being estimated. In these cases, computations on density functions (or likelihood functions) are necessary to achieve optimality [33].

Figure 2 shows a simple example in the context of cue integration. Changing the angle of a symmetric figure within its plane (its spin), keeping the 3D orientation of the plane itself fixed (imagine spinning a rectangle around a rod oriented perpendicular to the rectangle), changes the perceived 3D orientation of the plane, even when viewed in stereo [16]. These spin-dependent biases are well accounted for by a Bayesian model that optimally combines skew symmetry information (represented by a highly non-Gaussian likelihood function in Figure 2) with stereoscopic information about 3D surface orientation. The results would not be predicted by a deterministic scheme of weighting the estimates derived from each cue individually.

Perceptual biases and priors

Several recent studies on the role of prior models in perception provide strong evidence for Bayesian computations of the type envisioned here [15,18]. Weiss and Adelson's [17] recent work on motion provides an illustrative example. They propose a remarkably simple Bayesian model of motion perception in which likelihood functions derived from local, ambiguous motion measurements are multiplied together with a simple prior that biases interpretations to favor low speeds. The interplay between the likelihood functions and the prior density function leads to a complex pattern of directional biases that depends on contrast, edge orientation and other stimulus factors. The model predicts a surprisingly large range of previously unintuitive motion phenomena, suggesting that the brain could perform a similar computation.

Uncertainty and the control of action

The previous examples were all based on the performance of subjects in perceptual tasks. Sensory information, however, primarily serves the function of guiding action in the world. Researchers have recently begun developing techniques for coupling optimal Bayesian estimators to

control systems to maximize performance in motor tasks [34]. In much the same way that sensory uncertainty determines the optimal weighting scheme for combining sensory cues for perceptual judgments, sensory and motor uncertainties determine how sensory signals should be used to plan and control movements.

Consider the problem of using sensory feedback from the hand to guide online corrections of hand movements. Because of noise in the motor system and in initial sensory estimates of hand and target position, the movements of an individual are never perfect. Visual feedback from the moving hand should, in theory, be used to make small adjustments to movement trajectories online. How much an observer should trust the visual feedback, however, depends on how reliable it is. For example, when the hand is in the periphery of the visual field, visual estimates of position will necessarily be worse than when it is in the foveal area of the field (and near the target). Similarly, the error in motion signals from the hand scales with velocity, so that motion signals are least reliable around the point of peak velocity. Recent psychophysical studies have shown that humans use continuous feedback from the hand to control pointing movements, but that the relative contributions of different signals (e.g. position and velocity) depend on the expected sensory noise associated with those signals. Moreover, when noise is artificially added to visual feedback about the position of the hand, subjects optimally adjust the degree to which they rely on the feedback to make corrections [14,35,36].

Likewise, how one plans movements should depend on the intrinsic variability in the motor output and the costs associated with various errors. Recent behavioral tests have confirmed that motor plans take into account the uncertainty in motor outputs: ballistic movement trajectories effectively minimize the error in pointing movements, given the signal-dependent properties of motor noise [8], and when costs and gains for different aim points are independently varied, subjects adjust their aim points for fast pointing movements to maximize the expected gain [19,20]. Although not definitive, these results suggest that the brain uses knowledge of the uncertainty in the sensory input and the motor output for visuomotor control.

Neural representations of uncertainty

The notion that neural computations take into account the uncertainty of the sensory and motor variables raises two important questions: (i) how do neurons, or rather populations of neurons, represent uncertainty, and (ii) what is the neural basis of statistical inferences? Several schemes have been proposed over the past few years, which we now briefly review.

Binary variables

The simplest schemes apply to binary variables – that is, variables that can take only two states. This situation arises when subjects are asked to decide between two possibilities, such as whether an object is moving up or down. In this case, the uncertainty of the variable can be encoded in two ways. First, one can use two populations of neurons: one in which neurons respond proportionally to the probability of upward motion, and one in which they
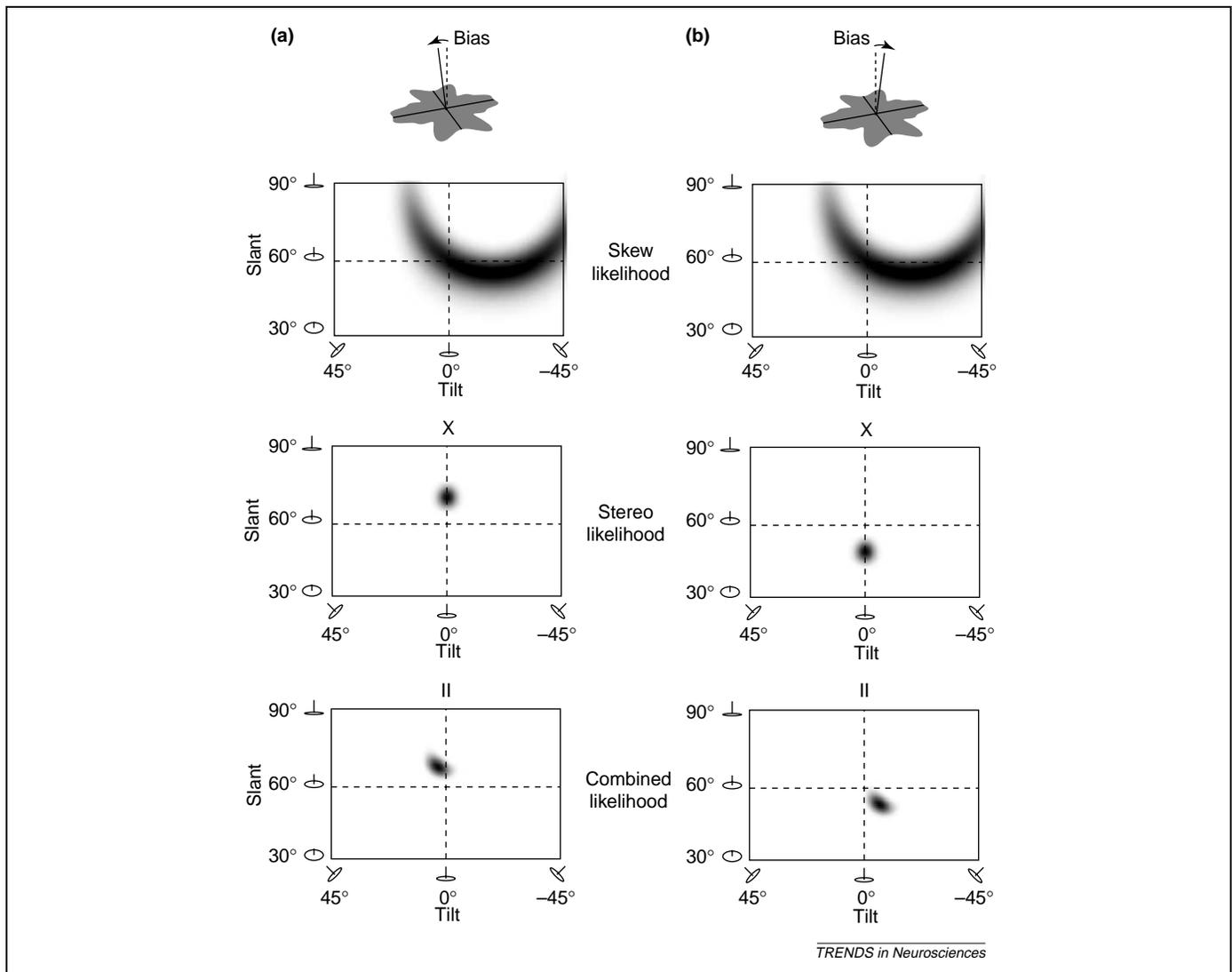
Figure 2. Skew symmetrical figures appear as figures slanted in depth because the brain assumes that they are projected from bilaterally symmetrical figures in the world. The information provided by skew symmetry is given by the angle between the projected symmetry axes of a figure, shown at the top of each panel as solid lines superimposed on the figure. Assuming that visual measurements of the orientations of these angles in the image are corrupted by Gaussian noise, one can compute a likelihood function for 3D surface orientation from skew. The result, as shown on the top two graphs, is highly non-Gaussian (darker points indicate higher likelihood values). The shape of the likelihood function is highly dependent on the spin of the figure around its 3D surface normal. When combined with stereoscopic information from binocular disparities (middle), an optimal estimator multiplies the likelihood functions associated with skew and stereo, as indicated by the 'X' between the top and middle panels, to produce a posterior distribution for surface orientation (bottom) given both cues (assuming the prior on surface orientation is flat). This leads to the prediction that perceptual biases will depend on the spin of the figure. It also leads to the somewhat counterintuitive prediction illustrated here that changing the slant suggested by stereo disparities should change the perceived tilt of symmetric figures. **(a)** When binocular disparities suggest a slant greater than that from which a figure was projected, the posterior distribution is shifted away from the orientation suggested by the disparities in both slant and tilt, creating a biased percept of the tilt of the figure. **(b)** The same figure, when binocular disparities suggest a smaller slant, gives rise to a tilt bias in the opposite direction. This is exactly the pattern of behavior shown by subjects.

respond to the probability of downward motion. An even simpler scheme involves only one population in which neurons respond proportionally to the ratio of upward and downward motion, equivalent to a quantity known as a likelihood ratio. Single-cell recordings in the lateral intraparietal (LIP) area provide evidence for both schemes. When monkeys are trained to perform one of two possible saccades, a subset of LIP neurons respond proportionally to the probability that the saccade ends in their receptive field [37]. However, when monkeys are trained to distinguish between two possible directions of motion, some LIP neurons integrate information over time in a way consistent with the computation of a likelihood ratio [38]. Similar ideas have also been suggested to interpret neuronal responses in the superior colliculus [39].

Convolution codes and variations

When direction of motion is not confined to two choices, but can take any value around a circle, the previous schemes no longer work. Rather, the brain needs an encoding mechanism that can deal with continuous variables. As described previously, the most natural way to represent uncertainty in a continuous variable is to represent the probability density function of the variable (Figure 3a). Any probability density function, $p(x)$, can be represented by its value at a few sample points along the x axis. The more samples are used, the better the approximation. This is the idea behind convolution codes, except that convolution codes avoid gaps between the samples by filtering $p(x)$ with a Gaussian kernel before sampling [40,41]. Each neuron simply
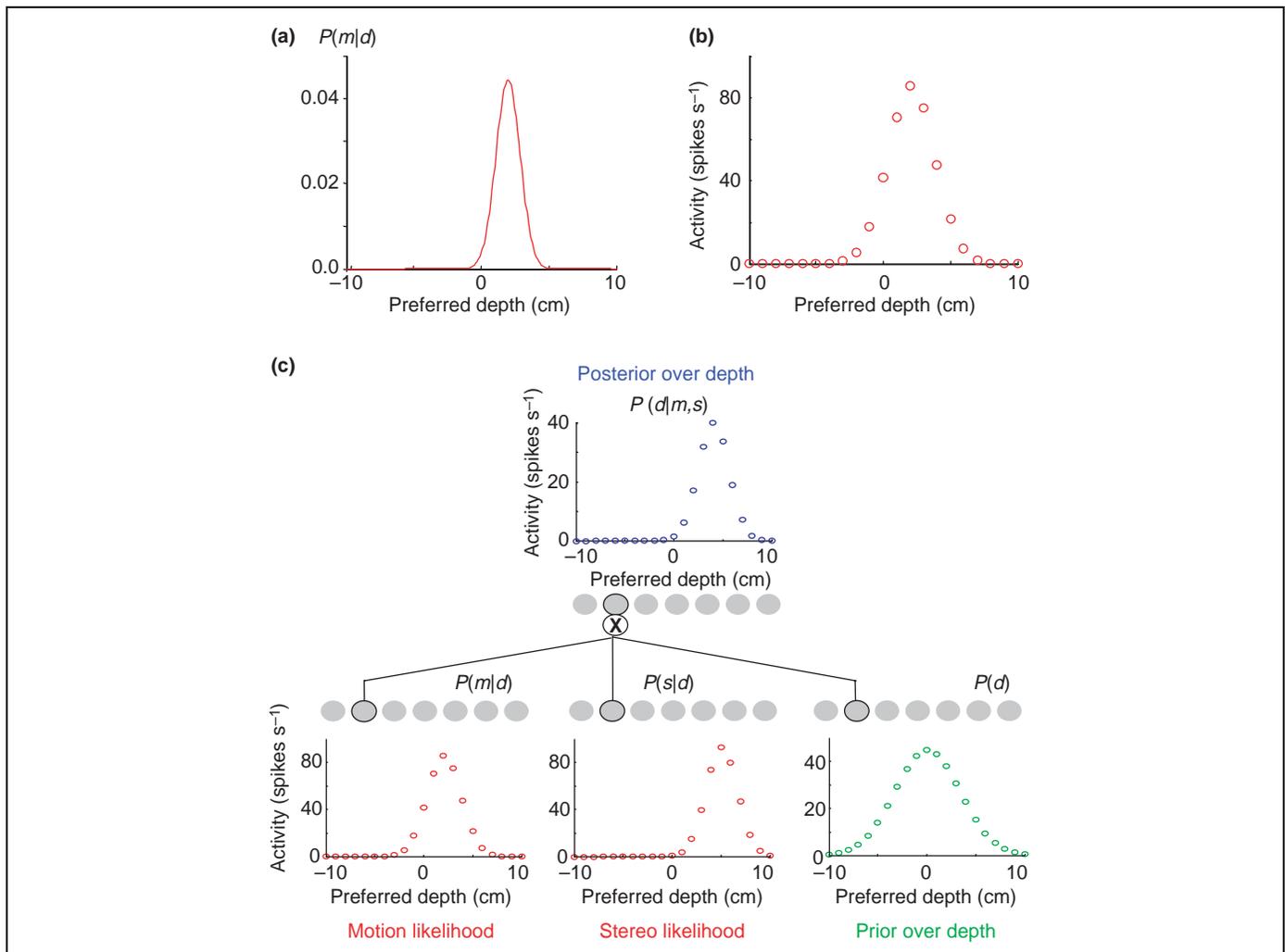
Figure 3. Inferences with convolution codes. **(a)** A hypothetical probability density function over motion given the depth of an object, $P(m|d)$. **(b)** A convolution code for the probability density function in (a). Each dot corresponds to the activity of one neuron plotted at its preferred depth. The code is basically a set of samples of the probability density function filtered, or convolved, by a Gaussian kernel. The width of this pattern of activity is related to the width of the function and, therefore, to the uncertainty of the encoded variable. **(c)** Given two likelihood functions $P(m|d)$ and $P(s|d)$, and one prior distribution $P(d)$, the posterior distribution over depth $P(d|m,s)$, is obtained by taking a neuron-by-neuron product of the three distributions (neurons are indicated by shaded circles, ranked by their preferred depth). This scheme works only if the Gaussian kernel used in the convolution code is a Dirac function, but it can easily be adapted to Gaussian kernels of any width [43]. Adapted from Ref. [46] © (2003) Annual Reviews (www.annualreviews.org).

computes the dot product between its Gaussian tuning curve and the probability density function. Assuming that the tuning curves of the neurons are just translated copies of the same Gaussian profile, the resulting population pattern of activity looks like the original probability density function filtered by the Gaussian tuning curve.

Inferences with convolution codes are straightforward. For instance, given motion and stereo disparity cues for depth, one would like to compute the posterior density function for depth conditioned on motion and stereo information by multiplying the likelihood functions for the observed motion of an object given its depth $p(m|d)$ with the observed pattern of horizontal disparities given its depth $p(s|d)$ and a prior distribution over depth. If neurons represent samples of the likelihood functions and the prior density function, a simple point-by-point product operation between the two representations (Figure 3c) is equivalent to multiplying the functions themselves [30,42]}. In reality, this scheme works only when the tuning curves of the neurons are Dirac functions (infinitely narrow Gaussian tuning curves) but it can be easily extended to

wider tuning curves [42,43]. This scheme has also been applied to static and time-varying variables [40,44].

Another possibility for representing a probability density function with a population code is to encode the log of the probability density function, instead of the function itself [45]. With this scheme, the point-by-point product required for Bayesian inference with convolution codes (Figure 3c) is replaced by a simple point-by-point sum, because logs turn products into sums. Using sums is appealing because of its simplicity and because it is consistent with the way LIP neurons integrate (i.e. sum) evidence over time [38].

Gain encoding
An alternative to the convolution code is to use what we call a gain-encoding scheme [46,47]. This scheme takes advantage of the near-Poisson nature of neural noise [48] to code the mean and variance of a density function simultaneously. To understand how the scheme works, consider the example of orientation selectivity. The primary visual cortex contains neurons with bell-shaped

tuning curves for orientation [49]. If we rank the neurons by their preferred orientations, the population response to a trial of particular orientation θ_0 takes the form of a hill of activity (Figure 4b). On any given trial, the shape of the hill is corrupted by near-Poisson noise. To decode such noisy population codes, one can use a Bayesian decoder which returns the posterior distribution over θ given the hill of activity, $p(\theta|A)$ [50,51]. For independent Poisson noise, the posterior distribution is Gaussian, with its mean controlled mostly by the position of the peak of the hill and the variance inversely proportional to the gain of the hill [46]. This is because, for Poisson noise, the variance of the spike count is proportional to the gain. This implies that the signal-to-noise ratio – the ratio of the gain over the square root of the variance – grows with the square root of the gain. Therefore, a high gain entails a high signal-to-noise ratio, and a narrow posterior distribution. Consequently, the noisy hill of activity can be treated as a neural code for the posterior, with the position of the peak encoding the mean, and the amplitude (or gain) encoding the variance.

Deneve *et al.* [47] have designed a network architecture that uses gain-encoding to perform optimal Bayesian inferences. They applied their network to the problem of locating an object based on its sound and image. On each trial, the network is initialized with two noisy population codes for the position of an object in visual and auditory coordinates (Figure 5). It then performs two tasks. First, it remaps the visual input into auditory coordinates and vice versa, through a basis function layer. This is a prerequisite for combining these signals because the visual system encodes the location of the object in retinal coordinates whereas the auditory system uses head-centered coordinates. The basis function units act as the building blocks of the transformation: they compute Gaussian functions of the visual and auditory input that are combined to approximate both changes of coordinates, just as a set of cubes can be combined to approximate any three dimensional shape. Second, the network recovers the maximum likelihood estimate of the object position given the visual and auditory inputs (Figure 5). This computation is the result of a relaxation process that turns the noisy population codes into smooth hills of activity over time. For a particular value of the network parameters, these smooth hills of activity peak very close to the maximum likelihood estimate

of the position. The cues are integrated with weights proportional to their reliability because noisy hills with high gain – corresponding to more reliable cues – provide a stronger initial push and, as a result, have a stronger influence of the final state of the network [47,52].

Although originally applied to object localization, this architecture can be generalized to any cue integration problem. In particular, this approach can be used to account for the performance of human observers in the experiments on cue integration [26–32]. The model can also be extended to time varying problems, such as estimating the position of a moving arm [4].

The gain-encoding model suggests a particularly intriguing role for Poisson variability. At first sight, it would appear that this variability is highly detrimental and severely limits the accuracy with which cortical circuits perform computations. The gain-encoding idea suggests that Poisson noise might in fact be very beneficial: it allows population codes to represent the mean as well as the variance of the encoded variables, the latter being crucial for Bayesian inferences.

It is important to emphasize that the different encoding schemes we have reviewed are not mutually exclusive. Uncertainties can take many forms; for example, the uncertainty due to photon noise in the retina has little to do with the ambiguity due to the aperture problem in motion processing. It is therefore possible that the brain uses multiple encoding schemes. Ultimately, which schemes are used in the brain can be answered only empirically. It is our hope that the accumulation of behavioral data showing that neural computation is akin to a Bayesian inference, and the development of several models of Bayesian inference in neural networks, will compel neurophysiologists to design experiments to test the predictions of these models.

Discussion

We have described psychophysical evidence that shows human observers to behave in a variety of ways like optimal Bayesian observers. The most compelling features of these data in regard to the Bayesian coding hypothesis are: (i) that subjects implicitly 'adjust' cue weights in a Bayes' optimal way based on stimulus and viewing parameters; (ii) that perceptual and motor behavior reflect a system that takes into account the uncertainty of both sensory and motor signals; (iii) that humans behave
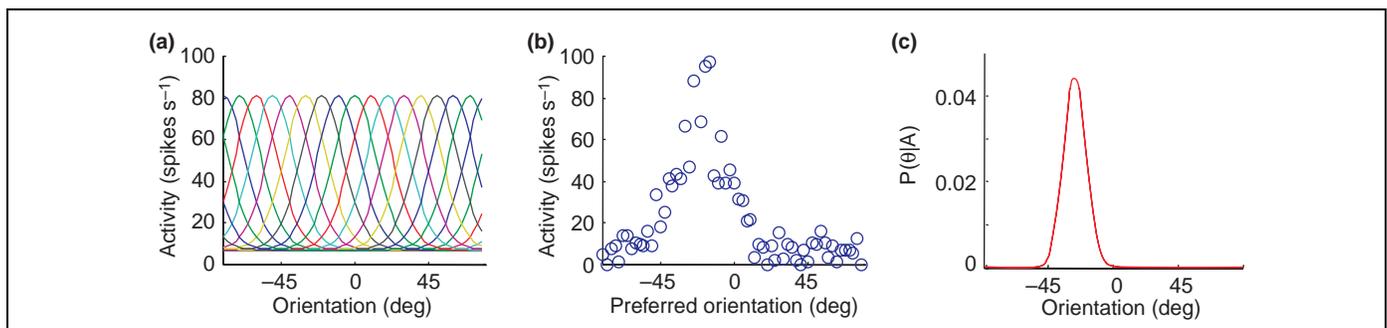

Figure 4. Inferences with gain encoding. **(a)** Idealized Gaussian tuning curves to orientation for 16 cells in primary visual cortex. **(b)** Response of 64 cells with Gaussian tuning curves similar to the ones shown in (a), in response to an orientation of $-20°$. The cells have been ranked according to their preferred orientation and the responses have been corrupted by independent Poisson noise, a good approximation of the noise observed *in vivo*. **(c)** The posterior distribution over orientation obtained from applying a Bayesian decoder to the noisy hills shown in (b). With independent Poisson noise, the peak of the distribution is given by the peak of the noisy hill, and the width of the distribution (i.e. the uncertainty) is inversely proportional the amplitude of the noisy hill. Adapted from Ref. [47].
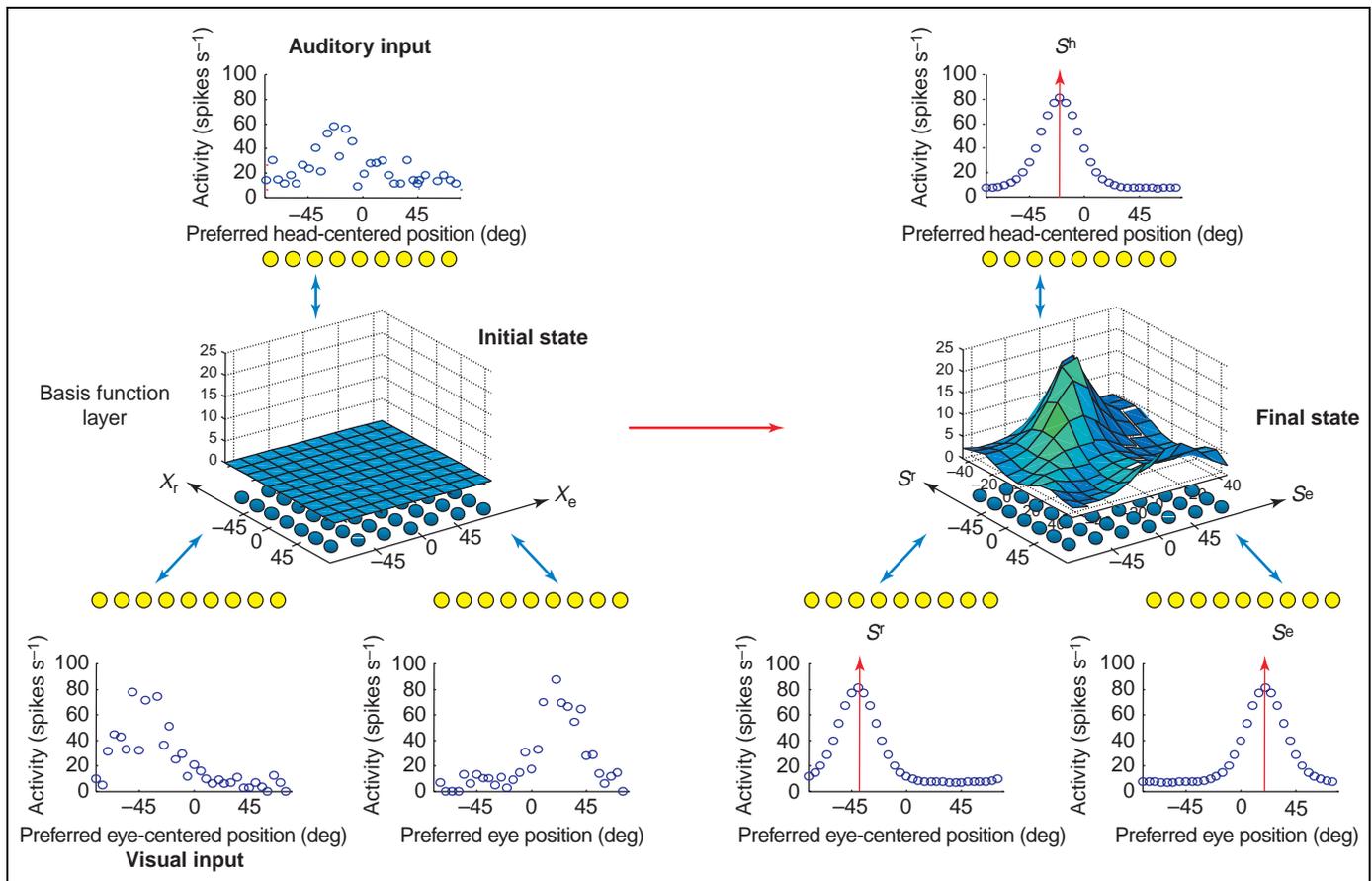
Figure 5. A basis function network for optimal cue integration between a visual and an auditory input. The network is tuned to perform two tasks simultaneously. Through the basis function layer, eye-centered visual inputs (X_r) are remapped into auditory coordinates (head-centered) and vice versa (this requires knowledge of the position of the eyes, which is encoded in a third population code, X_e). Over time, the network settles onto three smooth hills of activity, peaking in the final state at the location of the maximum likelihood estimates of the eye-centered position (S^r) and the head-centered position (S^h) of the target and at the position of the eyes in the head (S^e). This solution takes into account the respective reliability of each of the cues, as expected for a maximum likelihood estimate. In the case illustrated here, the auditory input is less reliable than the visual one, as indicated by the fact that the auditory hill has a reduced gain (top left). Adapted from Ref. [47].

near-optimally even when the sensory information is characterized by highly non-Gaussian density functions, leading to complex patterns of predicted behavior; and (iv) that relatively simple Bayesian models can account for otherwise complex patterns of subjective, perceptual biases, even when the prior density functions built into the models do not explicitly code the biases. We argue that these data strongly suggest that the brain codes even complex patterns of sensory and motor uncertainty in its internal representations and computations. Two challenges for further research emerge from this review.

First, although a growing body of psychophysical work is being devoted to exploring the ways in which humans are optimal observers and actors, an equally important challenge for future work is to find the ways in which human observers are not optimal. Owing to the complexity of the tasks, unconstrained Bayesian inference is not a viable solution for computation in the brain. Recent work on statistical learning, for example, has elucidated strong limits on the types of statistical regularities that sensory systems automatically detect [53–55].

Second, neuroscientists must begin to test theories of how uncertainty could be represented in populations of neurons. We have described several neural coding strategies that might be used to encode probability density

functions or their statistical moments; however, the neurophysiological evidence for these schemes is weak. This is not because existing data conflict with the strategies but, rather, because little work has been done to test them. Pursuing this challenge will require further development and application of advanced multi-electrode recording techniques. As these techniques mature, we hope that neuroscientists will take up the challenge to submit the Bayesian coding hypothesis to rigorous falsification tests.

Finally, we acknowledge that the examples presented here are much simpler than many of the perceptual estimation problems faced by the brain. Most notable among the complexities of these problems is their high dimensionality (e.g. contour completion and flow field estimation). Until recently, the problem of efficiently representing and computing with probability density functions in high-dimensional spaces has been a barrier to developing efficient Bayesian computer vision algorithms. The recent development of graphical models [22] and particle filtering techniques [56,57] has shown the most promise for implementing efficient Bayesian algorithms for high-dimensional estimation problems. How these might be implemented by the brain is a major challenge for computational neuroscientists [58].

References

1 Mach, E. (1980) *Contributions to the Analysis of the Sensations (C. M. Williams, Trans.)*, Open Court Publishing Co., Chicago, IL USA

2 Helmholtz, H. (1925) *Physiological Optics, Vol. III: The perceptions of Vision (J. P. Southall, Trans.)*, Optical Society of America, Rochester, NY USA

3 Liu, Z. *et al*. (1995) Object classification for human and ideal observers. *Vision Res.* 35, 549–568

4 Wolpert, D.M. *et al*. (1995) An internal model for sensorimotor integration. *Science* 269, 1880–1882

5 Crowell, J.A. and Banks, M.S. (1996) Ideal observer for heading judgments. *Vision Res.* 36, 471–490

6 Eagle, R.A. and Blake, A. (1995) 2-Dimensional constraints on 3-dimensional structure-from-motion tasks. *Vision Res.* 35, 2927–2941

7 Knill, D.C. (1998) Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Res.* 38, 1683–1711

8 Harris, C.M. and Wolpert, D.M. (1998) Signal-dependent noise determines motor planning. *Nature* 394, 780–784

9 van Beers, R.J. *et al*. (2001) Sensorimotor integration compensates for visual localization errors during smooth pursuit eye movements. *J. Neurophysiol.* 85, 1914–1922

10 van Beers, R.J. *et al*. (2002) Role of uncertainty in sensorimotor control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1137–1145

11 Shimozaki, S.S. *et al*. (2003) An ideal observer with channels versus feature-independent processing of spatial frequency and orientation in visual search performance. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 2197–2215

12 Beutter, B.R. *et al*. (2003) Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1341–1355

13 Murray, R.F. *et al*. (2003) Saccadic and perceptual. performance in visual search tasks. II. Letter discrimination. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1356–1370

14 Saunders, J.A. and Knill, D.C. (2004) Visual feedback control of hand movements. *J. Neurosci.* 24, 3223–3234

15 Mamassian, P. and Landy, M.S. (2001) Interaction of visual prior constraints. *Vision Res.* 41, 2653–2668

16 Saunders, J.A. and Knill, D.C. (2001) Perception of 3D surface orientation from skew symmetry. *Vision Res.* 41, 3163–3183

17 Weiss, Y. *et al*. (2002) Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604

18 van Ee, R. *et al*. (2003) Bayesian modeling of cue interaction: bistability in stereoscopic slant perception. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1398–1406

19 Trommershauser, J. *et al*. (2003) Statistical decision theory and trade-offs in the control of motor response. *Spat. Vis.* 16, 255–275

20 Trommershauser, J. *et al*. (2003) Statistical decision theory and the selection of rapid, goal-directed movements. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1419–1433

21 Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo CA, USA

22 Weiss, Y. and Freeman, W.T. (2001) Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Comput.* 13, 2173–2200

23 Freeman, W.T. *et al*. (2000) Learning low-level vision. *Int. J. Comput. Vis.* 40, 25–47

24 Geisler, W.S. and Diehl, R.L. (2003) A Bayesian approach to the evolution of perceptual and cognitive systems. *Cogn. Sci.* 27, 379–402

25 Geisler, W.S. (1989) Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96, 267–314

26 Jacobs, R.A. (1999) Optimal integration of texture and motion cues to depth. *Vision Res.* 39, 3621–3629

27 Knill, D.C. and Saunders, J.A. (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* 43, 2539–2558

28 Hillis, J.M. *et al*. Slant from texture and disparity cues: optional cue combination. *J. Vis.* (in press)

29 van Beers, R.J. *et al*. (1999) Integration of proprioceptive and visual position-information: an experimentally supported model. *J. Neurophysiol.* 81, 1355–1364

30 Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433

31 Battaglia, P.W. *et al*. (2003) Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1391–1397

32 Alais, D. and Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262

33 Knill, D.C. (2003) Mixture models and the probabilistic structure of depth cues. *Vision Res.* 43, 831–854

34 Todorov, E. and Jordan, M.I. (2002) Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5, 1226–1235

35 Saunders, J.A. and Knill, D.C. (2003) Humans use continuous visual feedback from the hand to control fast reaching movements. *Exp. Brain Res.* 152, 341–352

36 Kording, K.P. and Wolpert, D.M. (2004) Bayesian integration in sensorimotor learning. *Nature* 427, 244–247

37 Platt, M.L. and Glimcher, P.W. (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238

38 Gold, J.I. and Shadlen, M.N. (2001) Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5, 10–16

39 Anastasio, T.J. *et al*. (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Comput.* 12, 1165–1187

40 Anderson, C. (1994) Neurobiological computational systems. In *Computational Intelligence Imitating Life*, pp. 213–222, IEEE Press, New York NY, USA

41 Zemel, R. *et al*. (1998) Probabilistic interpretation of population code. *Neural Comput.* 10, 403–430

42 Zemel, R.S. and Dayan, P. (1997) Combining probabilistic population codes. In *JCAI-97: 15th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Francisco CA, USA

43 Barber, M.J. *et al*. (2003) Neural representation of probabilistic information. *Neural Comput.* 15, 1843–1864

44 Eliasmith, C. and Anderson, C.H. (2003) *Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems*, MIT Press

45 Rao, R.P. (2004) Bayesian computation in recurrent neural circuits. *Neural Comput.* 16, 1–38

46 Pouget, A. *et al*. (2003) Inference and computation with population codes. *Annu. Rev. Neurosci.* 26, 381–410

47 Deneve, S. *et al*. (2001) Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* 4, 826–831

48 Tolhurst, D. *et al*. (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* 23, 775–785

49 Hubel, D. and Wiesel, T. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154

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

51 Foldiak (1993) The 'ideal homunculus': statistical inference from neural population responses. In *Computation and Neural Systems* (Eeckman, F. and Bower, J. eds), pp. 55–60, Kluwer Academic Publishers

52 Latham, P.E. *et al*. (2003) Optimal computation with attractor networks. *J. Physiol. Paris* 97, 683–694

53 Fiser, J. and Aslin, R.N. (2002) Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15822–15826

54 Fiser, J. and Aslin, R.N. (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467

55 Fiser, J. and Aslin, R.N. (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504

56 Blake, A. *et al*. (1998) Statistical models of visual shape and motion. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* 356, 1283–1301

57 Isard, M. and Blake, A. (1998) Condensation: conditional density propagation for visual tracking. *Int. J. Comput. Vis.* 29, 5–28

58 Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448