Review

Organizing probabilistic models of perception

Wei Ji Ma

Department of Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, Houston TX 77030, USA

Probability has played a central role in models of perception for more than a century, but a look at probabilistic concepts in the literature raises many questions. Is being Bayesian the same as being optimal? Are recent Bayesian models fundamentally different from classic signal detection theory models? Do findings of near-optimal inference provide evidence that neurons compute with probability distributions? This review aims to disentangle these concepts and to classify empirical evidence accordingly.

Decision-making in an uncertain world

In order to survive and thrive, all animals must derive knowledge about the world from sensory observations. A wildebeest needs to know whether a predator is hiding in the high grass, a badminton player where the shuttlecock will land, and an actuary how long a life insurance buyer will live. As is typical of perception and cognition, each of these judgments is made under uncertainty, caused by noise and ambiguity in the observations [1]. Since Helmholtz [2] and Mach [3], scientists have used the language of probability to describe human perception under uncertainty. Signal detection theory [4–6] opened a new era in psychology, as it allowed researchers to model a wide range of tasks using the idea that observers make optimal decisions based on uncertain information, a strategy that involves Bayes' rule for probabilities. In recent decades, interest in Bayesian models of perception has surged, with many studies concluding that humans use information about sensory uncertainty in perceptual decision-making [7,8]. An appealing aspect of Bayesian models is that they are often [9], though not always [10], highly constrained by the statistical structure of the experiment and thus require few assumptions. Unfortunately, amidst the enthusiasm for probabilistic models of perception, important distinctions between the key notions of optimal inference, Bayesian inference, and computing with probability distributions are sometimes lost. This has led to inaccurate claims about perception and misplaced criticism of Bayesian models. The purpose of this review is to clarify the differences between these concepts. Although I will focus on perception, much of the discussion applies to other forms of cognition, as well.

Formalizing optimality, Bayesian inference, and probabilistic computation

An observer's knowledge of a world state can mathematically be expressed as a probability distribution – in the examples above, over target presence, landing location, and life span. Since this knowledge is based on sensory observations, the probability distribution is a conditional distribution, which can be denoted by $q(\text{world state} \mid \text{observations})$.

Knowledge is not sufficient for organisms; actions are needed. The wildebeest might decide whether to stay put, the badminton player whether to attempt a return, and the actuary what premium to set. Cost or utility is associated with each combination of true world state and action, denoted by $C(\text{world state, action})$: if the badminton player does not attempt to return the shuttle, energy is saved, but at the cost of a point if the shuttle lands inside the court. For the observer, the expected cost of an action is a weighted average over world states, with weights given by the probabilities of those world states given the observations:

$$EC(\text{action}) = \sum_{\text{world state}} C(\text{world state, action}) \times q(\text{world state}|\text{observations}) \quad (1)$$

Observers are called optimal if their actions minimize expected cost. Thus, optimality is defined only with respect to a specific cost function. The literature on cost functions used by human and non-human observers is extensive and has been reviewed elsewhere [8,11]. Although real-life tasks involve complex and often unknown cost functions, for simplicity, I will focus here on purely perceptual tasks, in which the observer's action is merely a report of the world state. In such tasks, it is reasonable to assume that the observer maximizes accuracy; this is sometimes reinforced by the observer receiving a unit reward for each correct report.

Regardless of the cost function, an observer who wants to achieve optimality must compute the conditional distribution $q(\text{world state} \mid \text{observations})$ on a given trial in a specific way, namely by using the statistics of world state and observations across many trials. Imagine that you are trying to guess the age of a man you just met. Different ages occur with different frequencies in the population, which can be expressed using a probability distribution $p(\text{age})$. The observations could consist of the man's visual image. Since the same age can give rise to an infinite number of different images, there is no one-to-one correspondence between world state and observations. Moreover, external and internal noise will add variability to the observations. Thus, the observations are best described by a conditional distribution $p(\text{visual observations} \mid \text{age})$. The distributions $p(\text{age})$ and $p(\text{visual observations} \mid \text{age})$ together make up the statistical structure of the task, also called the generative model [12,13].

Corresponding author: Ma, W.J. (wjma@bcm.edu)

1364-6613/$ – see front matter © 2012 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.tics.2012.08.010 Trends in Cognitive Sciences, October 2012, Vol. 16, No. 10

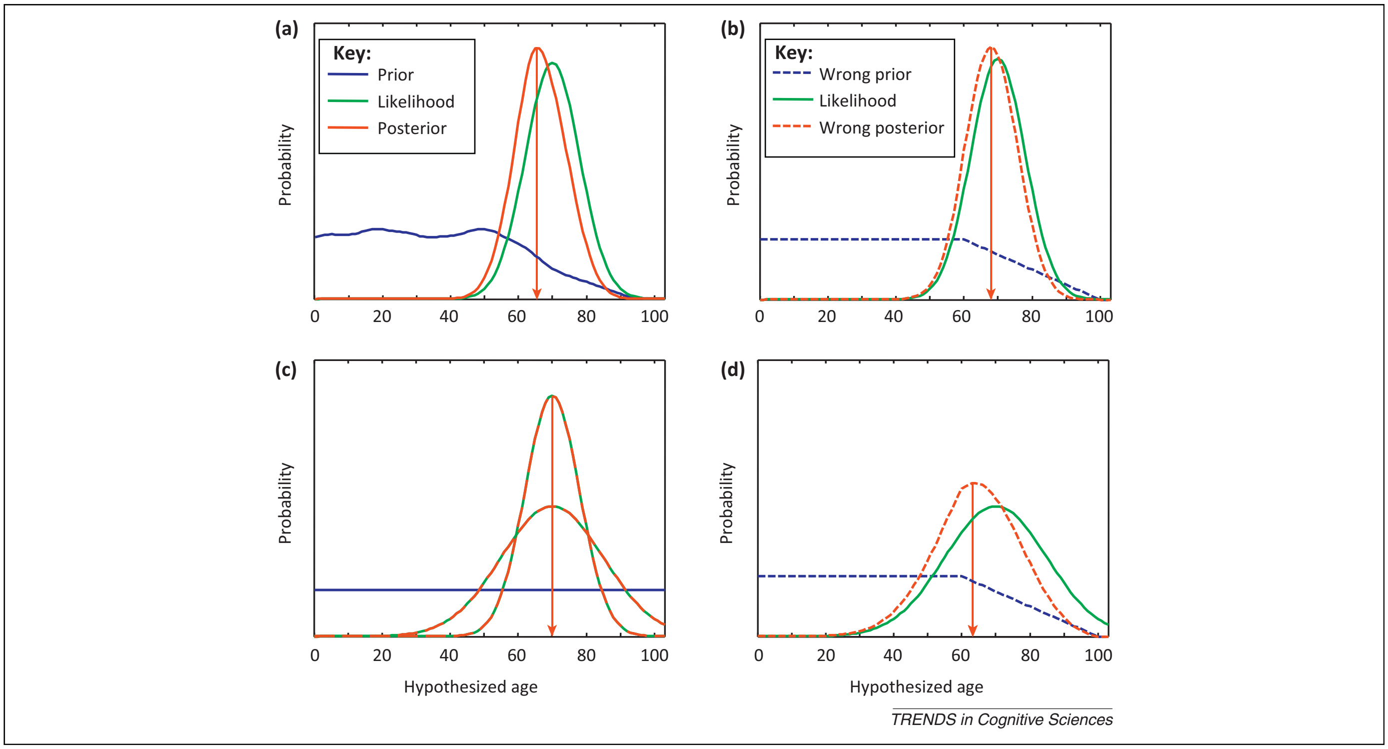

Figure 1. Distinguishing Bayesian, optimal, and probabilistic computation. **(a)** Bayesian inference of a man's age. The prior was obtained from the age distribution of men in the United States. (U.S. Census Bureau, 2010 Census Briefs: Age and Sex Composition 2010). The red arrow indicates the MAP estimate. **(b)** Using an approximate prior produces a different posterior and, though Bayesian, is suboptimal. The degree of suboptimality depends on the severity of the approximation. **(c)** If both the true age distribution and the observer's prior were flat, the MAP estimate would be independent of uncertainty (the width of the likelihood function; two examples shown) – non-probabilistic, yet optimal computation. **(d)** Even when inference is suboptimal, the MAP estimate may depend on uncertainty: here, the likelihood is broader than in **(b)** and the MAP estimate is different. This is an example of probabilistic, suboptimal computation.

An observer who knows the distributions in the generative model can compute the desired probability distribution over the world state – called the posterior distribution – using Bayes' rule (Figure 1a):

$$p(\text{world state}|\text{observations}) \propto p(\text{observation}|\text{world state}) \times p(\text{world state}) \quad (2)$$

All three probabilities in this equation are 'subjective': they quantify the observer's internal beliefs on a given trial, rather than the frequencies of outcomes across many trials. 'World state' should thus be interpreted as 'hypothesized world state'. In this interpretation, p(world state) quantifies the belief in each hypothesis in the absence of any observations and is called the prior distribution, whereas p(observations | world state) measures the strength of the evidence for each hypothesis and is called the likelihood function (of the world state). Using the posterior distribution from equation (2) for the distribution q(world state | observations) in equation (1) is necessary for minimizing expected cost. The computation in equation (2) is called Bayesian inference. To maximize accuracy (which corresponds to the cost function we assume for perception), the observer should pick the world state with the highest posterior probability. This is called maximum-a-posteriori (MAP) estimation.

Key dichotomies

The different concepts in probabilistic models of perception are listed in Table 1. This section describes and clarifies the key dichotomies in such models.

Optimal versus suboptimal Bayesian inference

Optimal inference and Bayesian inference are not synonymous, because Bayesian MAP estimation based on incorrect assumptions about the generative model is usually suboptimal. In the age estimation example, suppose you immigrated from a country with a more skewed age distribution, such that $q(\text{age}) \neq p(\text{age})$, or where men develop wrinkles at a later age, so that $q(\text{visual observations} \mid \text{age}) \neq p(\text{visual observations} \mid \text{age})$. It would then be perfectly Bayesian to use q(age) and q(visual observations | age) to compute posteriors in your new country; however, you would make more mistakes and thus not be optimal (Figure 1b). In perception, suboptimal Bayesian inference occurs in two important circumstances. First, the stimulus statistics in an experiment might not match those encountered in the natural world. Training notwithstanding, observers might use 'default' priors and likelihoods obtained from the statistics of the natural world. This effect is cleverly exploited in experiments such as the McGurk-MacDonald effect [14], where a small conflict is introduced between an auditory and a visual speech cue, presented simultaneously. The natural prior favors that synchronous audio and video originate from the same speech, but is incorrect in the experiment. Many cue conflict studies [7,15], in which subjects are presented with two slightly disparate stimuli but are still expected to combine them into a single percept, test Bayesian models that are suboptimal with respect to the experimental statistics. However, since natural statistics are undoubtedly more fundamental than experimental ones, an observer's use of natural statistics in inference is

Table 1. Concepts in probabilistic models of perception

Bayesian inference	Making a decision about a state of the world based on sensory observations by computing a posterior distribution. Can be optimal or suboptimal, and probabilistic or not.
Optimal inference	Making a decision about a state of the world based on sensory observations in such a way that expected cost is minimized. Is always Bayesian. Equivalent to MAP estimation when accuracy is maximized.
Probabilistic computation	Making a decision about a state of the world based on sensory observations, while taking into account knowledge of the uncertainty associated with the observations. Requires a trial-to-trial neural representation of uncertainty.
Absolute optimality	See optimal inference, with the relevant observations being the raw sensory input
Relative optimality	See optimal inference, with the relevant observation being a processed representation of the sensory input
Signal detection theory	Theory of how observers make decisions based on sensory observations. In principle very general, but in practice mostly restricted to binary decisions and non-probabilistic computation.

widely regarded as optimal. Second, suboptimal Bayesian inference can occur when the true distributions p(world state) and/or p(observations | world state) are difficult to learn or cannot be implemented in neural circuitry. In these situations, the brain might be forced to use approximations q(world state) and/or q(observations | world state), and again be suboptimal. Characterizing such approximations is an important direction for future research (Box 1). In summary, all optimal inference is Bayesian, but certainly not all Bayesian inference is optimal.

Probabilistic versus non-probabilistic computation

If an observer's behavior is found to be consistent with MAP estimation and thus with optimality, it is tempting to conclude, based on equation (2), that the observer encodes and computes with probability distributions. This conclusion would be premature. Even though modelers compute the MAP estimate from the posterior distribution, the MAP estimate is ultimately nothing but a specific function of the observations – a type of stimulus-response mapping. As long as the observer applies this function, behavior will be optimal, regardless of whether a posterior distribution is computed. To address the question of whether an observer encodes and computes with probability distributions, one can test models in which the observer's report on a given trial depends on the sensory uncertainty associated with each observation on that trial. If such a model (whether optimal or suboptimal) describes the observer's behavior well, one can conclude that at least some probabilistic information – namely sensory uncertainty – is propagated from the input stage to the decision. We call such computation *probabilistic*. Not all optimal computation is probabilistic (Figure 1c) and not all probabilistic computation is optimal (Figure 1d). As a consequence, the statement that 'Bayesian optimality implies that neurons encode probabilities' [16] is inaccurate. A correct statement would be 'If subjects use a decision rule that requires knowledge of sensory uncertainty, this implies that neurons encode uncertainty.'

The notion of probabilistic computation can help to distinguish signal detection theory models from the more recent wave of Bayesian models. For classic detection and discrimination tasks, signal detection theory models are optimal Bayesian models. Beyond such tasks, however, signal detection theory modelers have focused on non-probabilistic and sometimes non-Bayesian decision rules (examples are discussed below). By contrast, many recent Bayesian studies have examined tasks in which optimal performance does require computing with information about sensory uncertainty, that is, probabilistic computation.

Absolute versus relative optimality

So far, the nature of the observations has not been specified. The observations could consist of the physical signals entering the organism's sensory organs – the retina, hair cells, skin mechanoreceptors, etc. If the observations are understood in this way, the behavior of an optimal observer is optimal in an absolute sense: no system could possibly achieve a lower expected cost. Relative optimality is less stringent: the observations consist of an unspecified representation of the input inside the brain, which might contain less information than the raw sensory input.

Example: visual search

To make these notions concrete, let us consider a laboratory version of the wildebeest's conundrum – visual search. An observer reports whether an item oriented at 2° (the target) is present among N items; non-target items (distractors) all have a -2° orientation (Figure 2a). The probability that the target is present is 50%, and if it is present, it is equally likely to appear at any location. Let us denote the orientations on a given trial by $s_1,\ldots, s_N$, and the observer's noisy measurements (observations) by $x_1,\ldots, x_N$. In neural terms, x_i can be regarded as the best possible

Box 1. Future directions for Bayesian models of perception

- *Increasing complexity.* Most generative models of tasks in which models of probabilistic computation have been tested have been relatively simple. For example, most cue combination studies involve a single, one-dimensional stimulus variable. To examine how prevalent probabilistic computation is in perception, tasks with greater complexity have to be examined, such as categorization tasks and tasks with multiple objects. An ambitious direction is to apply Bayesian ideas to object recognition [12] and natural scene perception.
- *Modeling approximate, suboptimal inference.* Discovering the boundaries of optimality requires the development and testing of plausible suboptimal models, for instance ones inspired by machine learning [88]. The extent to which human behavior is optimal might depend on the complexity of the task's generative model and the task's ecological importance.
- *Resource constraints.* Biological constraints, such as on the number of available neurons and the amount of available energy [89,90], might affect decision-making, even in simple perceptual tasks. In psychological terms, resource constraints might take the form of spatial attention or limited working memory. In Bayesian models of multiple-object tasks, resource constraints have been modeled through a decrease in measurement precision with the number of items [36,76,91–95]. Resource constraints need to be characterized better and integrated with models of inference.
- *Neural implementation.* See Box 2.

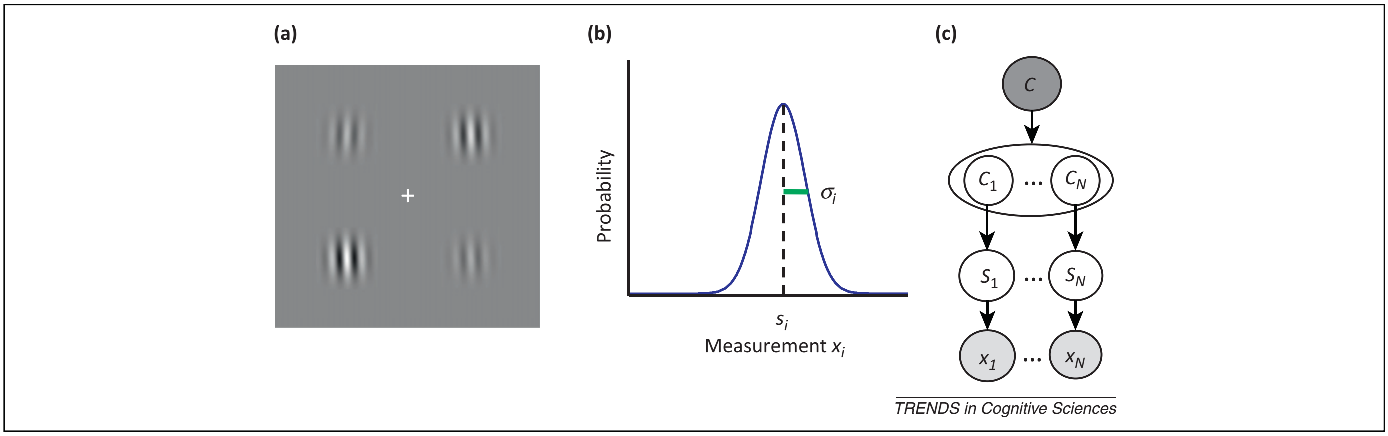

Figure 2. **(a)** A visual search task. The observer is asked to report whether a 2° rightward tilted line is present among distractors tilted 2° leftward. Items can have different contrasts. The display is shown briefly (e.g., for 100 ms). **(b)** We assume that the observer's noisy measurement x_i follows a Gaussian distribution around the true orientation s_i (here equal to 2° or −2°), with standard deviation σ_i (controlled by contrast). **(c)** Generative model. C denotes target presence (yes or no) in the whole display, C_i at the $i^{\text{th}}{}_i$ location. Each arrow represents a statistical dependence between variables. The observer infers C from the measurements $x_1,\ldots,x_N$. The generative model is completely determined by the experimental design combined with the assumption in **(b)**, and in turn completely determines the observer's decision rule and the Bayesian model's predictions for the observer's behavior.

guess of the orientation s_i based on single-trial neural activity. We make the standard assumption [6] that each x_i is drawn independently from a Gaussian distribution with mean s_i and standard deviation σ_i (Figure 2b). This standard deviation (noise level) can be experimentally manipulated through contrast. The statistical structure (generative model) of the task is shown graphically in Figure 2c. Let us now consider different decision rules, which can be organized according to Figure 3(I will refer to the numbered regions in this diagram; since the task is binary, Regions 2, 5, 8, and 11 of the diagram are not applicable.)

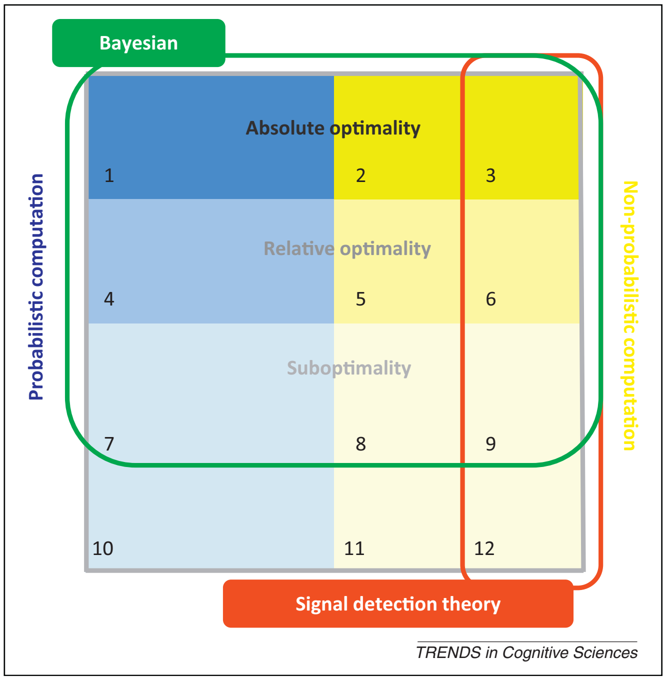

Figure 3. Probabilistic models of perception can be organized according to whether the observer maximizes reward (optimality) or whether the observer uses trial-to-trial uncertainty information (probabilistic computation). Optimality comes in two forms, absolute and relative. Bayesian and signal detection theory models occupy subsets of this space. Numbered regions are referenced in the text.

Regions 1 and 4: optimal, probabilistic computation

On a given trial, the observer has to decide whether the target is present based on a set of measurements $x_1,\ldots,x_N$. Using the generative model, one can derive, in a lengthy but straightforward manner, the decision rule an observer should apply to the measurements to maximize accuracy [5,17]. That optimal rule is used to report 'target present' when the quantity $\exp(4x_i/\sigma_i^2)$, averaged across locations, exceeds 1: $\frac{1}{N}\sum_{i=1}^{N} e^{\frac{4x_i}{\sigma_i^2}} > 1$. Every positive measurement x_i will contribute to evidence for the target being present, but importantly, every measurement is also weighted by its inverse variance, σ_i^2: noisier measurements are weighted less in the decision. If x_i and σ_i represent the information contained in the raw sensory signals, then this rule will produce absolute optimality (Region 1), otherwise relative optimality (Region 4).

From the point of view of the observer, σ_i is the level of sensory uncertainty about the i^{th} orientation. Thus, in order to be optimal in this task, the observer must know the level of sensory uncertainty associated with each measurement, and the decision rule is therefore an instance of probabilistic computation. Humans seem to follow this decision rule and thus perform both near-optimal and probabilistic computation in visual search [17].

Regions 3 and 6: optimal, non-probabilistic computation

When only a single item is presented (N=1), the search task reduces to a discrimination task (was the orientation −2° or 2°?) and the optimal decision rule above to $x_1>0$. This decision rule is non-probabilistic, since σ_1 does not appear. In simple detection and discrimination tasks, the equivalent of this rule is often a good description of human behavior [6], but this is of limited interest given the absence of plausible alternative models.

Regions 7 and 9: suboptimal, Bayesian computation

Bayesian inference under a wrong assumption about the generative model will usually produce a suboptimal decision rule. For example, if the observer erroneously believed that the target was present on 40% rather than 50% of trials, the decision rule would be $\frac{1}{N}\sum_{i=1}^{N} e^{\frac{4x_i}{\sigma_i^2}} > 1.5$, or if the observer believed that all measurements were equally noisy, the rule would be $\frac{1}{N}\sum_{i=1}^{N} e^{\frac{4x_i}{\sigma^2}} > 1$, with σ the assumed common noise level. A radically wrong belief about the generative model would be that on a target-present trial, all items are targets instead of only one. The Bayesian MAP rule under this belief would be $\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} > 0$, a variant of the so-called sum rule [18]. Each of these rules would lead to lower accuracy than the optimal rule. Each of these three suboptimal rules also happens to be probabilistic (Region 7). A non-probabilistic and suboptimal, but Bayesian decision rule would be the standard sum rule, $\sum_{i=1}^{N} x_i > 0$ [18], which is the MAP rule for a generative model in which on a target-present trial, all items are targets, and in addition, all items have the same noise level (Region 9). Neither the sum rule nor its probabilistic variant can describe human behavior [17].

Regions 10 and 12: suboptimal, non-Bayesian computation

A plausible, non-probabilistic, suboptimal decision rule is the maximum-of-outputs or max rule, $\max_i x_i > k$, where k is a constant criterion. This rule is not only suboptimal, but also non-Bayesian (Region 12), since it is not derived from MAP estimation based on a generative model. The max rule has had a long history in signal detection theory [18–20], but was recently shown to be inadequate when uncertainty varies unpredictably [17]. A probabilistic version of the max rule could be $\max_i \frac{x_i}{\sigma_i} > k$ (Region 10).

Psychophysical evidence

There is psychophysical support for each of the types of computation discussed above.

Relative optimality

A fertile testing ground for relative optimality has been the study of how humans and monkeys combine cues from different modalities, for example, visual and vestibular information about self-motion direction. Some cue combination studies find evidence for near-optimality [21–26], but others report marked deviations from optimality in the form of overweighting of one of the cues [27–29]. In the realm of evidence accumulation (cue combination over time), optimality is somewhat ambiguous in view of the speed-accuracy trade-off, but the prominent diffusion-to-bound model for binary choice [30,31] is closely related to a model that is optimal in a specific sense [32,33]. Demonstrations of near-optimal inference in tasks more complex than cue combination have emerged in recent years [17,25,34–39].

Several Bayesian studies have examined whether human subjects perform MAP estimation using priors derived from natural statistics [40]. For example, the statistics of contours in natural scenes predict whether subjects judge a set of oriented elements as belonging to the same contour [41–43], whereas the statistics of orientations are consistent with human priors in an orientation judgment task [44]. These studies provide evidence for optimal, non-probabilistic computation.

Absolute optimality

Absolute optimality, usually non-probabilistic, has been examined in many contexts, including photon detection [45–49], pattern detection and discrimination [50–53], discrimination of dot density [54], object recognition [55,56], and letter identification [57] (see [9] for a detailed review). Very few instances of absolute near-optimality have been reported, however. Efficiency (closeness to absolute optimality) seems task-dependent, potentially reflecting suboptimal inference rather than the mere presence of noise in the nervous system [58].

Bayesian inference without optimality

Priors over many variables, such as speed [59,60], temporal duration [61], event location [34], surface shape and orientation [62], and illuminant chromaticity in color perception [63] have been postulated within the context of Bayesian models. Since those priors were not derived from experimental or natural statistics, these studies provide evidence for Bayesian and incidentally also probabilistic computation, but not for optimality. A recent contour detection study found that humans performed worse than an absolutely optimal observer who uses the experimental statistics, and argued for a Bayesian, suboptimal, non-probabilistic model [64]; an earlier study proposed a Bayesian, non-probabilistic model, but did not address optimality [65].

Probabilistic computation

A strong test of probabilistic computation is to vary sensory noise unpredictably from trial to trial and (if applicable) from item to item, while withholding feedback or providing uninformative feedback. This has been done in several cue combination studies [22,23,26], visual search [17], confidence-based visual decisions [66], and change detection [39]. Weaker tests are those using only a single level of noise [24,34,67,68], because a subject could use multiple trials to estimate sensory uncertainty, and those providing trial-to-trial feedback [21,35], because a subject could, instead of using internal knowledge of sensory uncertainty, treat the corresponding variables as unknown parameters and learn their values through the feedback. In combining perceptual information with an asymmetric reward structure, both humans [68,69] and non-human animals [70–72] use an internal estimate of decision confidence, thus suggesting probabilistic, but not necessarily optimal computation. Explicit tests of probabilistic computation in evidence accumulation, by manipulating stimulus reliability over time, have been rare [73]. Finally, signal detection theory studies of multi-object tasks have mostly focused on testing max and sum models [18,19,74–79], that is, non-probabilistic computation. To my knowledge, no instances exist in which optimality requires probabilistic computation, but human data are better described by a model of non-probabilistic computation.

It should be noted that some published probabilistic models of perception have technical problems. A tempting mistake when sensory noise is present is to assume that the likelihood function, such as the noise distribution, is fixed for a given stimulus (e.g., [80–84]). In reality, the likelihood function and, therefore, the posterior vary from trial to trial even when the stimulus is kept fixed, because they are computed from the noisy measurements on each trial.

Concluding remarks

In this article, I drew distinctions between the notions of Bayesian inference, optimality, and probabilistic computation, and motivated a corresponding organization of empirical evidence. Psychophysical evidence shows that absolute optimality is elusive, paints a mixed picture regarding relative optimality, but provides strong indications that the brain performs probabilistic computation and thus computes with neural representations of uncertainty. It is likely that in most real-world perception, organisms perform computations that are probabilistic, but suboptimal due to the complexity of generative models and limitations of neural circuitry. However, more tests of probabilistic computation are needed (Box 1).

In practice, high-level cognition is more difficult to model than perception, since cost functions and generative models are less constrained and more ambiguous. In principle, however, many of the distinctions drawn here also apply to other areas of cognition. For example, judging a man's total lifespan from a single observation of age can be formulated as a Bayesian inference problem [85]. Indeed, in the cited study, subjects were found to behave (as a group) close to optimally. However, this does not demonstrate that humans compute with likelihood functions or probability distributions; much more work is needed in this area.

Finally, psychophysical evidence for probabilistic computation raises the question how neurons compute with sensory uncertainty information or even full probability distributions. One scheme that has been proposed involves making use of the likelihood function naturally encoded in neural population activity (Box 2). Beyond perception, the neural implementation of cognitive probabilistic models remains virtually unexplored [86,87].

Box 2. Neural models of probabilistic computation

Models of perception have the potential to constrain neural implementations of perceptual computations. To form a neural model out of a behavioral model, the first step is to define the relevant level of neural variables. A strong candidate is the level of spike counts in sensory and decision-making neurons [96]. For example, presenting an oriented stimulus s might elicit a set of spike counts $\mathrm{r}=(r_1,\ldots,r_n)$ in a population of orientation-tuned cells in primary visual cortex. Population activity is subject to trial-to-trial variability, which can be described by a distribution $p(\mathrm{r}|s)$. There is a direct connection between r and x, the scalar measurement of the stimulus in a behavioral model: the latter is the value that maximizes the neural likelihood function over the stimulus, $L(s)=p(\mathrm{r}|s)$ [97,98]. The width of the likelihood function $L(s)$ is the observer's uncertainty about the stimulus, σ. Since r is a high-dimensional variable, it has sufficient degrees of freedom to encode σ on a trial-by-trial basis. For example, if we assume for $p(\mathrm{r}|s)$ independent Poisson variability with Gaussian tuning curves, then σ^2 is proportional to the inverse of the total spike count in the population [99]. Using neural likelihood functions, one can map Bayesian models of behavior to neural operations. This approach, also called probabilistic population coding, has been successfully applied to cue combination [16], decision-making [100,101], visual search [17], and various forms of marginalization (averaging over confounding variables) [102]. Many alternative schemes for encoding likelihood functions or probability distributions using neurons have been proposed [13,103–112].

References

1 Smith, A.M. (ed.) (2001) *Alhacen's Theory of Visual Perception: A Critical Edition, with English Translation and Commentary, of the First Three Books of Alhacen's De aspectibus, the Medieval Latin Version of Ibn al-Haytham's Kitab al-Manazir*, American Philosophical Society
2 Helmholtz, H.L.F.v. (1856) *Treatise on Physiological Optics*, Thoemmes Press
3 Mach, E. (1886/1897) *Contributions to the Analysis of the Sensations*, Open Court Publishing
4 Tanner, W.P. *et al.* (1954) A decision-making theory of visual detection. *Psychol. Rev.* 61, 401–409
5 Peterson, W.W. *et al.* (1954) The theory of signal detectability, In *Transactions IRE Profession Group on Information Theory, PGIT-4*, 171–212
6 Green, D.M. and Swets, J.A. (1966) *Signal Detection Theory and Psychophysics*, John Wiley & Sons
7 Knill, D.C. and Richards, W., eds (1996) *Perception as Bayesian Inference*, Cambridge University Press
8 Kording, K. (2007) Decision theory: what 'should' the nervous system do? *Science* 318, 606–610
9 Geisler, W.S. (2011) Contributions of ideal observer theory to vision research. *Vision Res.* 51, 771–781
10 Bowers, J.S. and Davis, C.J. (2012) Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414
11 Trommershauser, J. *et al.* (2008) Decision making, movement planning and statistical decision theory. *Trends Cogn. Sci.* 12, 291–297
12 Kersten, D. *et al.* (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304
13 Vilares, I. and Kording, K.P. (2011) Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Ann. N. Y. Acad. Sci.* 1224, 22–39
14 McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264, 746–748
15 Trommershauser, J. *et al.*, eds (2011) *Sensory Cue Integration*, Oxford University Press
16 Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438
17 Ma, W.J. *et al.* (2011) Behavior and neural basis of near-optimal visual search. *Nat. Neurosci.* 14, 783–790
18 Graham, N. *et al.* (1987) Signal dection models for multidimensional stimuli: probability distributions and combination rules. *J. Math. Psychol.* 31, 366–409
19 Nolte, L.W. and Jaarsma, D. (1966) More on the detection of one of *M* orthogonal signals. *J. Acoust. Soc. Am.* 41, 497–505
20 Verghese, P. (2001) Visual search and attention: a signal detection theory approach. *Neuron* 31, 523–535
21 Gu, Y. *et al.* (2008) Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* 11, 1201–1210
22 Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433
23 Alais, D. and Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262
24 van Beers, R.J. *et al.* (1996) How humans combine simultaneous proprioceptive and visual position information. *Exp. Brain Res.* 111, 253–261
25 Knill, D.C. (2003) Mixture models and the probabilistic structure of depth cues. *Vision Res.* 43, 831–854
26 Landy, M.S. and Kojima, H. (2001) Ideal cue combination for localizing texture-defined edges. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 18, 2307–2320
27 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
28 Fetsch, C.R. *et al.* (2012) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci.* 15, 146–154

29 Rosas, P. *et al.* (2005) Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 22, 801–809
30 Ratcliff, R. (1978) A theory of memory retrieval. *Psychological Review* 85, 59–108
31 Gold, J.I. and Shadlen, M.N. (2001) Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5, 10–16
32 Bogacz, R. *et al.* (2006) The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* 113, 700–765
33 Wald, A. and Wolfowitz, J. (1948) Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* 19, 326–339
34 Kording, K.P. *et al.* (2007) Causal inference in multisensory perception. *PLoS ONE* 2, e943
35 Van den Berg, R. *et al.* (2011) Optimal inference of sameness. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3178–3183
36 Vul, E. *et al.* (2009) Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Neural Inf. Proc. Syst.* 22, 1955
37 Sato, Y. *et al.* (2007) Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.* 19, 3335–3355
38 Hospedales, T. and Vijayakumar, S. (2009) Multisensory oddity detection as Bayesian inference. *PLoS ONE* 4, e4205
39 Keshvari, S. *et al.* (2012) Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE* 7, e40216
40 Geisler, W.S. (2008) Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167–192
41 Geisler, W.S. *et al.* (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* 41, 711–724
42 Geisler, W.S. and Perry, J.S. (2009) Contour statistics in natural images: Grouping across occlusions. *Vis. Neurosci.* 26, 109–121
43 Elder, J.H. and Goldberg, R.M. (2002) Ecological statistics of Gestalt laws for the perceptual organization of contours. *J. Vis.* 2, 324–353
44 Girshick, A.R. *et al.* (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14, 926–932
45 Barlow, H. (1956) Retinal noise and absolute threshold. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 46, 634–639
46 Cohn, T.E. and Lasley, D.J. (1974) Detectability of a luminance increment: effect of spatial uncertainty. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 64, 1715–1719
47 De Vries, H.D. (1943) The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica* 10, 553–564
48 Rose, A. (1948) The sensitivity performance of the human eye on an absolute scale. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 38, 196–208
49 Hecht, S. *et al.* (1942) Energy, quanta, and vision. *J. Gen. Physiol.* 25, 819–840
50 Burgess, A.E. *et al.* (1981) Efficiency of human visual signal discrimination. *Science* 214, 93–94
51 Burgess, A.E. and Colborne, B. (1988) Visual signal detection. IV. Observer inconsistency. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 5, 617–627
52 Legge, G.E. *et al.* (1987) Contrast discrimination in noise. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 4, 391–404
53 Geisler, W.S. (1989) Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96, 267–314
54 Barlow, H.B. (1978) The efficiency of detecting changes of density in random dot patterns. *Vision Res.* 18, 637–650
55 Tjan, B.S. *et al.* (1995) Human efficiency for recognizing 3-D objects in luminance noise. *Vision Res.* 35, 3053–3069
56 Liu, Z. *et al.* (1995) Object classification for human and ideal observer. *Vision Res.* 35, 549–568
57 Beckmann, P.J. and Legge, G.E. (2002) Preneural limitations on letter identification in central and peripheral vision. *J. Opt. Soc. Am. A: Opt. Image Sci. Vis.* 19, 2349–2362
58 Beck, J.M. *et al.* (2012) Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* 74, 30–39
59 Stocker, A.A. and Simoncelli, E.P. (2006) Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9, 578–585
60 Weiss, Y. *et al.* (2002) Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604
61 Ahrens, M. and Sahani, M. (2011) Observers exploit stochastic models of sensory change to help judge the passage of time. *Curr. Biol.* 21, 1–7
62 Mamassian, P. and Landy, M.S. (1998) Observer biases in the 3D interpretation of line drawings. *Vision Res.* 38, 2817–2832
63 Brainard, D.H. *et al.* (2006) Bayesian model of human color constancy. *J. Vis.* 6, 1267–1281
64 Ernst, U.A. *et al.* (2012) Optimality of human contour detection. *PLoS Comp. Biol.* 8, e1002520
65 Feldman, J. (2001) Bayesian contour integration. *Percept. Psychophys.* 63, 1171–1182
66 Barthelmé, S. and Mamassian, P. (2010) Flexible mechanisms underlie the evaluation of visual confidence. *Proc. Natl. Acad. Sci. U. S. A.* 107, 20834–20839
67 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
68 Whiteley, L. and Sahani, M. (2008) Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *J. Vis.* 8, 1–15
69 Gao, J.T. *et al.* (2010) Dynamic integration of reward and stimulus information in perceptual decision making. *PLoS ONE* 6, e16749
70 Kepecs, A. *et al.* (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–233
71 Kiani, R. and Shadlen, M.N. (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324, 759–764
72 Feng, S. *et al.* (2009) Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Comp. Biol.* 5, e1000284
73 Huk, A.C. and Shadlen, M.N. (2005) Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J. Neurosci.* 25, 10420–10436
74 Baldassi, S. and Verghese, P. (2002) Comparing integration rules in visual search. *J. Vis.* 2, 559–570
75 Eckstein, M.P. (1998) The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychol. Sci.* 9, 111–118
76 Wilken, P. and Ma, W.J. (2004) A detection theory account of change detection. *J. Vis.* 4, 1120–1135
77 Baldassi, S. and Burr, D. (2006) Visual clutter causes high-magnitude errors. *PLoS Biol.* 4, e56
78 Eckstein, M.P. *et al.* (2000) A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Percept. Psychophys.* 62, 425–451
79 Pelli, D.G. (1985) Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A: Opt. Image Proc. Syst.* 2, 1508–1532
80 Shams, L. *et al.* (2005) Sound-induced flash illusion as an optimal percept. *Neuroreport* 16, 1923–1927
81 Roach, N.W. *et al.* (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc. Biol. Sci.* 273, 2159–2168
82 Maij, F. *et al.* (2011) Temporal uncertainty separates flashes from their background during saccades. *J. Neurosci.* 31, 3708–3711
83 Brenner, E. *et al.* (2008) If I saw it, it probably wasn't far from where I was looking. *J. Vis.* 8, 1–10
84 Rowland, B. *et al.* (2007) A Bayesian model unifies multisensory spatial localization with the physiological properties of the superior colliculus. *Exp. Brain Res.* 180, 153–161
85 Griffiths, T. and Tenenbaum, J. (2006) Optimal predictions in everyday cognition. *Psychol. Sci.* 17, 767–773
86 Griffiths, T.L. *et al.* (2010) Probabilistic models of cognition: exploring representations and inductive biases. *Trends Cogn. Sci.* 14, 357–364
87 Shi, L. *et al.* (2010) Exemplar models as a mechanism for performing Bayesian inference. *Psychon. Bull. Rev.* 17, 443–464
88 Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*, Springer
89 Attwell, D. and Laughlin, S.B. (2001) An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145

90 Lennie, P. (2003) The cost of cortical computation. *Curr. Biol.* 13, 493–497
91 Bays, P.M. and Husain, M. (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854
92 Van den Berg, R. *et al.* (2012) Variability in encoding precision accounts for visual short-term memory limitations. *Proc. Natl. Acad. Sci. U.S.A.* http://dx.doi.org/10.1073/pnas.1117465109
93 Ma, W.J. and Huang, W. (2009) No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *J. Vis.* 9, 1–30
94 Shaw, M.L. (1980) Identifying attentional and decision-making components in information processing. In *Attention and Performance* (Nickerson, R.S., ed.), pp. 277–296, Erlbaum
95 Palmer, J. (1990) Attentional limits on the perception and memory of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 332–350
96 Barlow, H.B. (1972) Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394
97 Foldiak, P. (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
98 Sanger, T. (1996) Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.* 76, 2790–2793
99 Ma, W.J. (2010) Signal detection theory, uncertainty, and Poisson-like population codes. *Vision Res.* 50, 2308–2319
100 Beck, J.M. *et al.* (2008) Bayesian decision-making with probabilistic population codes. *Neuron* 60, 1142–1145
101 Huys, Q. *et al.* (2007) Fast population coding. *Neural Comput.* 19, 404–441
102 Beck, J.M. *et al.* (2011) Marginalization in neural circuits with divisive normalization. *J. Neurosci.* 31, 15310–15319
103 Fiser, J. *et al.* (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130
104 Anastasio, T.J. *et al.* (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Comput.* 12, 1165–1187
105 Deneve, S. (2008) Bayesian spiking neurons I: inference. *Neural Comput.* 20, 91–117
106 Anderson, C. (1994) Neurobiological computational systems. In *Computational Intelligence Imitating Life* (Zurada, J.M. *et al.*, eds), pp. 213–222, IEEE Press
107 Hoyer, P.O. and Hyvarinen, A. (2003) Interpreting neural response variability as Monte Carlo sampling of the posterior, In *Advances in Neural Information Processing Systems*, pp. 277–284, MIT Press
108 Barlow, H.B. (1969) Pattern recognition and the responses of sensory neurons. *Ann. N. Y. Acad. Sci.* 156, 872–881
109 Zemel, R. *et al.* (1998) Probabilistic interpretation of population code. *Neural Comput.* 10, 403–430
110 Rao, R.P. (2004) Bayesian computation in recurrent neural circuits. *Neural Comput.* 16, 1–38
111 Berkes, P. *et al.* (2011) Spontaneous cortical activity reveals hallmarks of an optimal internel model of the environment. *Science* 331, 83–87
112 Jazayeri, M. and Movshon, J.A. (2006) Optimal representation of sensory information by neural populations. *Nat. Neurosci.* 9, 690–696