



Published in final edited form as:

*Trends Cogn Sci.* 2010 March ; 14(3): 119–130. doi:10.1016/j.tics.2010.01.003.

## Statistically optimal perception and learning: from behavior to neural representations:

Perceptual learning, motor learning, and automaticity

József Fiser<sup>1,2</sup>, Pietro Berkes<sup>1</sup>, Gergő Orbán<sup>1,3</sup>, and Máté Lengyel<sup>4</sup>

<sup>1</sup> National Volen Center for Complex Systems, Brandeis University, Volen 208/MS 013, Waltham, MA 02454, USA

<sup>2</sup> Department of Psychology and the Neuroscience Program, Brandeis University, 415 South Street, Waltham, MA 02453, USA

<sup>3</sup> Department of Biophysics, Research Institute for Particle and Nuclear Physics, Hungarian Academy of Sciences, Konkoly Thege Miklós út 29–33, H-1121, Budapest, Hungary

<sup>4</sup> Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1 PZ, United Kingdom

### Abstract

Human perception has recently been characterized as statistical inference based on noisy and ambiguous sensory inputs. Moreover, suitable neural representations of uncertainty have been identified that could underlie such probabilistic computations. In this review, we argue that learning an internal model of the sensory environment is another key aspect of the same statistical inference procedure and thus perception and learning need to be treated jointly. We review evidence for statistically optimal learning in humans and animals, and reevaluate possible neural representations of uncertainty based on their potential to support statistically optimal learning. We propose that spontaneous activity can have a functional role in such representations leading to a new, sampling-based, framework of how the cortex represents information and uncertainty.

### Probabilistic perception, learning and representation of uncertainty: in need of a unifying approach

One of the longstanding computational principles in neuroscience is that the nervous system of animals and humans is adapted to the statistical properties of the environment [1]. This principle is reflected across all organizational levels, ranging from the activity of single neurons to networks and behavior, and it has been identified as key to the survival of organisms [2]. Such adaptation takes place on at least two distinct behaviorally relevant time scales: on the time scale of immediate inferences, as a moment-by-moment processing of sensory input (perception), and on a longer time scale by learning from experience. Although statistically optimal perception and learning have most often been considered in isolation, here we promote them as two facets of the same underlying principle and treat them together under a unified approach.

Although there is considerable behavioral evidence that humans and animals represent, infer and learn about the statistical properties of their environment efficiently [3], and there is also

converging theoretical and neurophysiological work on potential neural mechanisms of statistically optimal perception [4], there is a notable lack of convergence from physiological and theoretical studies explaining whether and how statistically optimal learning might occur in the brain. Moreover, there is a missing link between perception and learning: there exists virtually no crosstalk between these two lines of research focusing on common principles and on a unified framework down to the level of neural implementation. With recent advances in understanding the bases of probabilistic coding and the accumulating evidence supporting probabilistic computations in the cortex, it is now possible to take a closer look at both the basis of probabilistic learning and its relation to probabilistic perception.

We first provide a brief overview of the theoretical framework as well as behavioral and neural evidence for representing uncertainty in perceptual processes. To highlight the parallels between probabilistic perception and learning, we then revisit in more detail the same issues with regard to learning. We argue that a main challenge is to pinpoint representational schemes that enable neural circuits to represent uncertainty for both perception and learning, and compare and critically evaluate existing proposals for such representational schemes. Finally, we review a seemingly disparate set of findings regarding variability of evoked neural responses and spontaneous activity in the cortex and suggest that these phenomena can be interpreted as part of a representational framework that supports statistically optimal inference and learning.

## **Probabilistic perception: representing uncertainty, behavioral and neural evidence**

At the level of immediate processing, perception has long been characterized as unconscious inference, where incoming sensory stimuli are interpreted in terms of the objects and features that gave rise to them [5]. Traditional approaches treated perception as a series of classical signal processing operations, by which each sensory stimulus should give rise to a single perceptual interpretation [6]. However, because sensory input in general is noisy and ambiguous, there is usually a range of different possible interpretations compatible with any given input. A well-known example is the ambiguity caused by the formation of a two-dimensional image on our retina by objects that are three-dimensional in reality (Figure 1a). When such multiple interpretations arise, the mathematically appropriate way to describe them is to assign a probability value to each of them that expresses how much one believes that a particular interpretation might reflect the true state of the world [7], such as the true three-dimensional shape of an object in Figure 1a. Although the principles of probability theory have been established and applied to studying economic decision-making for centuries [8], only recently has their relevance to perception been appreciated, causing a quiet paradigm shift from signal processing to probabilistic inference as the appropriate theoretical framework for studying perception.

Evidence has been steadily growing in recent years that the nervous system represents its uncertainty about the true state of the world in a probabilistically appropriate way and uses such representations in two cognitively relevant domains: information fusion and perceptual decision-making. When information about the same object needs to be fused from several sources, inferences about the object should rely on these sources commensurate with their associated uncertainty. That is, more uncertain sources should be relied upon less (Figure 1b). Such probabilistically optimal fusion has been demonstrated in multisensory integration [9, 10] when the different sources are different sensory modalities, and also between information coming from the senses and being stored in memory [11,12].

Probabilistic representations are also key to decision-making under risk and uncertainty [13]. Making a well-informed decision requires knowledge about the true state of the world. When

there is uncertainty about the true state, then the decision expected to yield the most reward (or utility) can be computed by weighing the reward associated with each possible state of the world with their probabilities (Figure 1c). Indeed, several studies demonstrated that, in simple sensory and motor tasks, humans and animals do take into account their uncertainty in such a way [14].

The compelling evidence at the level of behavior that humans and animals represent uncertainty during perceptual processes initiated intense research into the neural underpinnings of such probabilistic representations. The two main questions that have been addressed is how sensory stimuli are represented in a probabilistic manner by neural cell populations (i.e. how neural activities encode probability distributions over possible states of the sensory world), and how the dynamics of neural circuits implement appropriate probabilistic inference with these representations [15,16]. As a result, there has been a recent surge of converging experimental and theoretical work on the neural bases of statistically optimal inference. This work has shown that the activity of groups of neurons in particular decision tasks can be related to probabilistic representations and that dynamical changes in neural activities are consistent with probabilistic computations with the represented variables [4,17].

In summary, the study of perception as probabilistic inference, for which the key is the representation of uncertainty, provides an exemplary synthesis of a sound theoretical background with behavioral as well as neural evidence.

## Probabilistic learning: representing uncertainty

In contrast to immediate processing during perception, the probabilistic approach to learning has been less explored in a neurobiological context. This is surprising given the fact that, from a computational standpoint, probabilistic inference based on sensory input is always made according to a model of the sensory environment which typically needs to be acquired by learning (Figure 2). Thus, the goal of probabilistic learning can be defined as acquiring appropriate models for inference based on past experience. Importantly, just as perception can be formalized as inferring hidden states, variables, of the environment from the limited information provided by sensory input (e.g. inferring the true three-dimensional shape and size of the seat of a chair from its two-dimensional projection on our retinae), learning can be formalized as inferring some more persistent hidden characteristics, parameters, of the environment based on limited experience. These inferences could target concrete physical parameters of objects, such as the typical height or width of a chair, or more abstract descriptors, such as the possible categories to which objects can belong (e.g. chairs and tables) (Figure 2).

There are two different ways in which representing uncertainty is important for learning. First, learning about our environment modifies the perceptual inferences we draw from a sensory stimulus. That is, the same stimulus gives rise to different uncertainties after learning. For example, having learned about the geometrical properties of chairs and tables allows us to increase our confidence that an unusually looking stool is really more of a chair than a table (Figure 2). At the neural level, this constrains learning mechanisms to change neural activity patterns such that they correctly encode the ensuing changes in perceptual uncertainty, thus keeping the neural representation of uncertainty self-consistent before and after learning. Second, representing uncertainty does not just constrain but also benefits learning. For example, if there is uncertainty as to whether an object is a chair or a table, our models for both of these categories should be updated, rather than only updating the model of the most probable category (Figure 2). Crucially, the optimal magnitude of these updates depends directly (and inversely) on the uncertainty about the object belonging to each category: the model of the more probable category should be updated to a larger degree [18].

Thus, probabilistic perception implies that learning must also be probabilistic in nature. Therefore, we now examine behavioral and neural evidence for probabilistic learning.

## Probabilistic learning: behavioral level

Evidence for humans and animals being sensitive to the probabilistic structure of the environment ranges from low-level perceptual mechanisms, such as visual grouping mechanisms conforming with the co-occurrence statistics of line edges in natural scenes [19], to high-level cognitive decisions such as humans' remarkably precise predictions about the expected life time of processes as diverse as cake baking or marriages [20]. A recent survey demonstrated how research in widely different areas ranging from classical forms of animal learning to human learning of sensorimotor tasks found evidence of probabilistic learning [21]. It has been found that configural learning in animals [22], causal learning in rats [23] as well as in human infants [24] and a vast array of inductive learning phenomena fit comfortably with a hierarchical probabilistic framework, in which probabilistic learning is performed at increasingly higher levels of abstraction [25].

A particularly direct line of evidence for humans learning complex, high-dimensional distributions of many variables by performing higher-order probabilistic learning, not just naïve frequency-based learning, comes from the domain of visual statistical learning (Box 1). An analysis of a series of visual statistical learning experiments showed that beyond the simplest results, recursive pairwise associative learning is inadequate for replicating human performance, whereas Bayesian probabilistic learning not only accurately replicates these results but it makes correct predictions about human performance in new experiments [26].

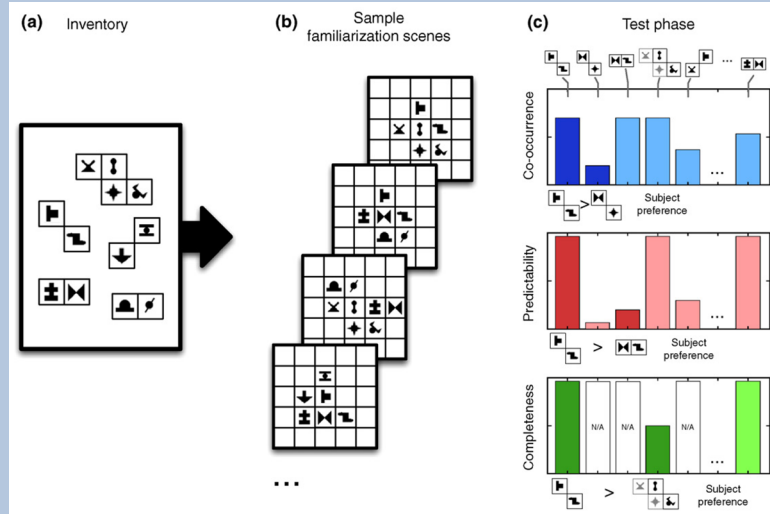
### Box 1

#### Visual statistical learning in humans

In experimental psychology, the term “statistical learning” refers to a particular type of implicit learning that emerged from investigating artificial grammar learning. It is fundamentally different from traditional perceptual learning and was first used in the domain of infant language acquisition [72]. The paradigm has been adapted from auditory to other sensory modalities such as touch and vision, to different species and various aspects of statistical learning have been explored such as multimodal interactions [73], effects of attention, interaction with abstract rule learning [74], together with its neural substrates [75]. The emerging consensus based on these studies is that statistical learning is a domain-general, fundamental learning ability of humans and animals that is probably a major component of the process by which internal representations of the environment are developed.

The basic idea of the statistical learning paradigm is to create an artificial mini-world by using a set of building blocks to generate several composite inputs that represent possible instances in this world. In the case of visual statistical learning (VSL), artificial visual scenes are composed from abstract shape elements where the building blocks are two or more such shapes always appearing in the same relative configuration (Figure I). An implicit learning paradigm is used to test how internal visual representations emerge through passively observing a large number of such composite scenes without any instruction as to what to pay attention to. After being exposed to the scenes, when subjects have to choose between two fragments of shape combinations based on familiarity, they reliably more often choose fragments that were true building blocks of the scenes compared to random combinations of shapes [76]. Similar results were found in 8-month-old infants [77,78] suggesting that humans from an early age can automatically extract the underlying structure of an unknown sensory data set based purely on statistical characteristics of the input.

Investigations of VSL provided evidence of increasingly sophisticated aspects of this learning, setting it apart from simple frequency-based naïve learning methods. Subjects not only automatically become sensitive to pairs of shapes that appear more frequently together, but also to pairs with elements that are more predictive of each other even when the co-occurrence of those elements is not particularly high [76]. Moreover, this learning highly depends on whether or not a pair of elements is a part of a larger building structure, such as a quadruple [79]. Thus, it appears that human statistical learning is a sophisticated mechanism that is not only superior to pairwise associative learning but also potentially capable to link appearance-based simple learning and higher-level “rule-learning” [26].



**Figure 1.**

Visual statistical learning. **(a)** An inventory of visual chunks is defined as a set of two or more spatially adjacent shapes always co-occurring in scenes. **(b)** Sample artificial scenes composed of multiple chunks that are used in the familiarization phase. Note that there are no obvious low-level segmentation cues giving away the identity of the underlying chunks. **(c)** During the test phase, subjects are shown pairs of segments that are either parts of chunks or random combinations (segments on the top). The three histograms show different statistical conditions. (Top) There is a difference in co-occurrence frequency of elements between the two choices; (middle) co-occurrence is equated, but there is difference in predictability (the probability of one symbol given that the other is present) between the choices; (bottom) both co-occurrence and predictability are equated between the two choices, but the completeness statistics (what percentage of a chunk in the inventory is covered by the choice fragment) is different – one pair is a standalone chunk, the other is a part of a larger chunk. Subjects were able to use cues in any of these conditions, as indicated by the subject preferences below each panel. These observations can be accounted for by optimal probabilistic learning, but not by simpler alternatives such as pairwise associative learning (see text).

These examples suggest a common core representational and learning strategy for animals and humans that shows remarkable statistical efficiency. However, such behavioral studies provide no insights as to how these strategies might be implemented in the neural circuitry of the cortex.

## Probabilistic learning in the cortex: neural level

Although psychophysical evidence has been steadily growing, there is little direct electrophysiological evidence showing that learning and development in neural systems is

optimal in a statistical sense even though the effect of learning on cortical representations has been investigated extensively [27,28]. One of the main reasons for this is that there have been very few plausible computational models proposed for a neural implementation of probabilistic learning that would provide easily testable predictions (but see Refs [29,30]). Here, we give a brief overview of the computational approaches developed to capture probabilistic learning in neural systems and discuss why they are unsuitable in their current form for being tested in electrophysiological experiments.

Classical work on connectionist models aimed at devising neural networks with simplified neural-like units that could learn about the regularities hidden in a stimulus ensemble [31]. This line of research has been developed further substantially and demonstrated explicitly how dynamical interactions between neurons in these networks correspond to computing probabilistic inferences, and how the tuning of synaptic weights corresponds to learning the parameters of a probabilistic model of input stimuli [32–34]. A key common feature in these statistical neural networks is that inference and learning are inseparable: inference relies on the synaptic weights encoding a useful probabilistic model of the environment, whereas learning proceeds by using the inferences produced by the network (Figure 3a).

Although statistical neural networks have considerably grown in sophistication and algorithmic efficiency in recent years [35], providing cutting-edge performance in some challenging real-world machine learning tasks, much less attention has been devoted to specifying their biological substrates. At the level of general insights, these models suggested ways in which internally generated spontaneous activity patterns (“fantasies”) that are representative of the probabilistic model encoded by the neural network can have important roles in the fine tuning of synapses during off-line periods of functioning. They also clarified that useful learning rules for such tuning always include Hebbian as well as anti-Hebbian terms [32,33]. In addition, several insightful ideas about the roles of bottom-up, recurrent, and top-down connections for efficient inference and learning have also been put forward [36–38], but they were not specified at a level that would allow direct experimental tests.

Learning internal models of natural images has traditionally been one area where the biological relevance of statistical neural networks was investigated. As these studies aimed at explaining the properties of early sensory areas, the “objects” they learned to infer were simple localized and oriented filters assumed to interact mostly additively in creating images (Figure 3b). Although these representations are at a lower level than the “true” objects constituting our environment (such as chairs and tables) that typically interact in highly non-linear ways as they form images (owing to e.g. occlusion [39]), the same principles of probabilistic inference and learning also apply to this level. Indeed, several studies showed how probabilistic learning of natural scene statistics leads to representations that are similar to those found in simple and complex cells of the visual cortex [40–44]. Although some early studies were not formulated originally in a statistical framework [40,41,43], later theoretical developments showed that their learning algorithms were in fact special cases of probabilistic learning [45,46].

The general method of validation in these learning studies almost exclusively concentrated on comparing the “receptive field” properties of model units with those of sensory cortical neurons and showing a good match between the two. However, as the emphasis in many of these models is on learning, the details of the mapping of neural dynamics to inference were left implicit (with some notable exceptions [44,47]). In cases where inference has been defined explicitly, neurons were usually assumed to represent single deterministic (so-called “*maximum a posteriori*”) estimates (Figure 3b). This failure to represent uncertainty is not only computationally harmful for inference, decision-making and learning (Figures 1–2) but it is also at odds with behavioral data showing that humans and animals are influenced by perceptual

uncertainty. Moreover, this approach constrains predictions to be made only about receptive fields which often says little about trial-by-trial, on-line neural responses [48].

In summary, presently a main challenge in probabilistic neural computation is to pinpoint representational schemes that enable neural networks to represent uncertainty in a physiologically testable manner. Specifically, learning with such representations on naturalistic input should provide verifiable predictions about the cortical implementation of these schemes beyond receptive fields.

## Probabilistic representations in the cortex for inference and learning

The conclusion of this review so far is that identifying the neural representation of uncertainty is key for understanding how the brain implements probabilistic inference and learning. Crucially, because inference and learning are inseparable, a viable candidate representational scheme should be suitable for both. In line with this, evidence is growing that perception and memory-based familiarity processes once thought to be linked to anatomically clearly segregated cortical modules along the ventral pathway of the visual cortex could rely on integrated multipurpose representations within all areas [49]. In this section, we review the two main classes of probabilistic representational schemes that are the best candidates for providing neural predictions and investigate their suitability for inference and learning.

Theoretical proposals of how probabilities can be represented in the brain fall into two main categories: schemes in which neural activities represent parameters of the probability distribution describing uncertainty in sensory variables, and schemes under which neurons represent the sensory variables themselves, such as models based on sampling. A simple example highlighting the core differences between the two approaches can be given by describing our probabilistic beliefs about the height of a chair (as in Figure 2). A parameter-based description starts with assuming that these beliefs can be described by one particular type of probability distribution, for example a Gaussian, and then specifies values for the relevant parameters of this distribution, for example the mean and (co)variance (describing our average prediction about the height of the chair and the “error bars” around it) (Figure 3c). A sampling-based description does not require that our beliefs can be described by one particular type of probability distribution. Rather, it specifies a series of possible values (samples) for the variable(s) of interest themselves, here the height of the particular chair viewed, such that if one constructed a histogram of these samples, this histogram would eventually trace out the probability distribution actually describing our beliefs (Figure 3c).

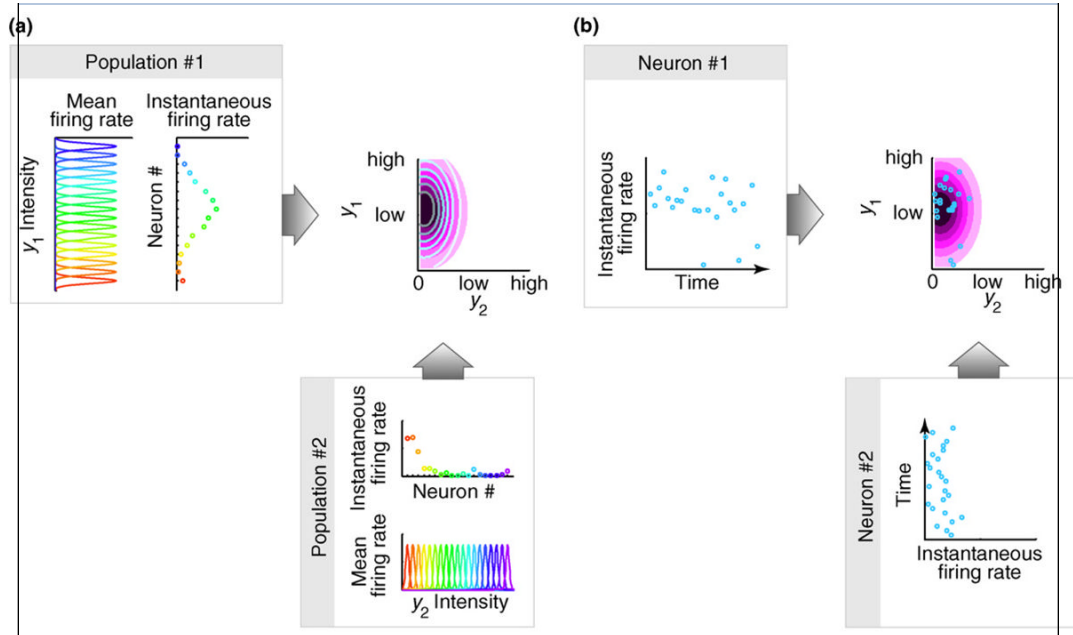
Probabilistic population codes (PPCs) are well-known examples of parameter-based descriptions and they are widely used for probabilistic modeling of inference making [16]. Recently, neurophysiological support for PPCs in cortical areas related to decision-making was also reported [50]. The key concept in PPCs, just as in their predecessors, kernel density estimator codes [51] and distributional population codes [15], is that neural activities encode parameters of the probability distribution that is the result of probabilistic inference (Box 2). As a consequence, a full probability distribution is represented at any moment in time and therefore changes in neural activities encode dynamically changing distributions as inferences are updated based on continuously incoming stimuli. Several recent studies explored how such representations can be used for optimal perceptual inference and decision-making in various tasks [52]. However, a main shortcoming of PPCs is that, at present, there is no clear option to implement learning within this framework.

**Box 2****Probabilistic representational schemes for inference and learning**

Representing uncertainty associated with sensory stimuli requires neurons to represent the probability distribution of the environmental variables that are being inferred. One class of schemes called probabilistic population codes (PPCs) assumes that neurons correspond to parameters of this distribution (Figure Ia). A simple but highly unfeasible version of this scheme would be if different neurons encoded the elements of the mean vector and covariance matrix of a multivariate Gaussian distribution. At any given time, the activities of neurons in PPCs provide a complete description of the distribution by determining its parameters, making PPCs and other parametric representational schemes particularly suitable for real-time inference [80–82]. Given that, in general, the number of parameters required to specify a multivariate distribution scales exponentially with the number of its variables, a drawback of such schemes is that the number of neurons needed in an exact PPC representation would be exponentially large and with fewer neurons the representation becomes approximate. Characteristics of the family of representable probability distributions by this scheme are determined by the characteristics of neural tuning curves and noisiness [16] (Table I).

An alternative scheme to represent probability distributions in neural activities is based on each neuron corresponding to one of the inferred variables. For example, each neuron can encode the value of one of the variables of a multivariate Gaussian distribution. In particular, the activity of a neuron at any time can represent a sample from the distribution of that variable and a “snapshot” of the activities of many neurons therefore can represent a sample from a high-dimensional distribution (Figure Ib). Such a representation requires time to take multiple samples (i.e. a sequence of firing rate measurements) for building up an increasingly reliable estimate of the represented distribution which might be prohibitive for on-line inference, but it does not require exponentially many neurons and—given enough time—it can represent any distribution (Table I). A further advantage of collecting samples is that marginalization, an important case of computing integrals that infamously plague practical Bayesian inference, learning and decision-making, becomes a straightforward neural operation. Finally, although it is unclear how probabilistic learning can be implemented with PPCs, sampling based representations seem particularly suitable for it (see main text).





**Figure I.**

Two approaches to neural representations of uncertainty in the cortex. **(a)** Probabilistic population codes rely on a population of neurons that are tuned to the same environmental variables with different tuning curves (populations 1 and 2, colored curves). At any moment in time, the instantaneous firing rates of these neurons (populations 1 and 2, colored circles) determine a probability distribution over the represented variables (top right panel, contour lines), which is an approximation of the true distribution that needs to be represented (purple colormap). In this example,  $y_1$  and  $y_2$ , are independent, but in principle, there could be a single population with neurons tuned to both  $y_1$  and  $y_2$ . However, such multivariate representations require exponentially more neurons (see text and Table I). **(b)** In a sampling based representation, single neurons, rather than populations of neurons, correspond to each variable. Variability of the activity of neurons 1 and 2 through time represents uncertainty in environmental variables. Correlations between the variables can be naturally represented by co-variability of neural activities, thus allowing the representation of arbitrarily shaped distributions.

The alternative, sampling-based approach to representing probability distributions is based on the idea that each neuron represents an individual variable from a high-dimensional multivariate distribution of external variables, and, therefore, each possible pattern of network activities corresponds to a point in this multivariate “feature” space (Box 2). Uncertainty is naturally encoded by network dynamics that stochastically explore a series of neural activity patterns such that the corresponding features are sampled according to the particular distribution that needs to be represented. Importantly, there exist worked out examples of how learning can be implemented in this framework: almost all the classical statistical neural networks have already been using this representational scheme implicitly [32,33,37]. The deterministic representations used in some of the most successful statistical receptive field models [42,43] can also be conceived as approximate versions of the sampling-based representational approach.

Despite its appeal for learning, there have been relatively few attempts to explicitly articulate the predictions of a sampling-based representational scheme for neural activities [53–55]. On

the level of general predictions, sampling models provide a natural explanation for neural variability and co-variability [56], as stochastic samples vary in order to represent uncertainty. They also provide an account of bistable perception, and its neural correlates [57]: multiple interpretations of the input correspond to multiple modes in the probability distribution over features; sequential sampling from such a distribution would produce in alternation samples from one of the peaks, but not from both at the same time [54,58]. Although to date there is no direct electrophysiological evidence reinforcing the idea that the cortex represents distributions through samples, sampling-based representations have been recently invoked to account for several psychophysical and behavioral phenomena including stochasticity in learning, language processing and decision-making [59–61].

Thus, although sampling-based representations are a promising alternative to PPCs, their neurobiological ramifications are much less explored at present. In the next section, we turn to spontaneous activity in the cortex, a phenomenon so far not discussed in the context of neural representation of uncertainty, and review its potential role in developing a sound theory of sampling-based models of probabilistic neural coding.

## Spontaneous activity and sampling-based representations

Modeling neural variability in evoked responses is an important first step in going beyond the modeling of receptive fields, and it is increasingly recognized as a critical benchmark for models of cortical functioning, including those positing probabilistic computations [16,48]. Another major challenge in this direction is to accommodate spontaneous activity recorded in the awake nervous system without specific stimulation (Box 3). From a signal processing standpoint, spontaneous activity has been long considered as nuisance elicited by various aspects of stochastic neural activity [62], even though some proposals exist that discuss potential benefits of noise in the nervous system [63]. However, several recent studies showed that the level of spontaneous activity is surprisingly high in some areas, and that it has a pattern highly similar to that of stimulus-evoked activity (Box 3). These findings suggest that a very large component of high spontaneous activity is probably not noise but might have a functional role in cortical computation [64,65].

### Box 3

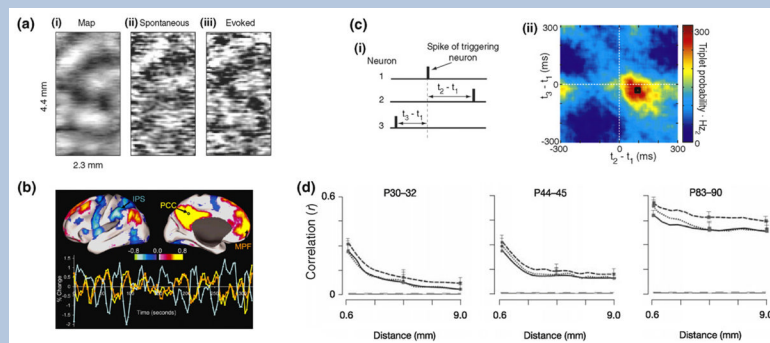
#### Spontaneous activity in the cortex

Spontaneous activity in the cortex is defined as ongoing neural activity in the absence of sensory stimulation [83]. This definition is the clearest in the case of primary sensory cortices where neural activity has traditionally been linked very closely to sensory input. Despite some early observations that it can influence behavior, cortical spontaneous activity has been considered stochastic noise [84]. The discovery of retinal and later cortical waves [85] of neural activity in the maturing nervous system has changed this view in developmental neuroscience, igniting an ongoing debate about the possible functional role of such spontaneous activity during development [86].

Several recent results based on the activities of neural populations initiated a similar shift in view about the role of spontaneous activity in the cortex during real-time perceptual processes [65]. Imaging and multi-electrode studies showed that spontaneous activity has large scale spatiotemporal structure over millimeters of the cortical surface, that the mean amplitude of this activity is comparable to that of evoked activity and it links distant cortical areas together [64,87,88] (Figure I). Given the high energy cost of cortical spike activity [89], these findings argue against the idea of spontaneous activity being mere noise. Further investigations found that spontaneous activity shows repetitive patterns [90,91], it reflects the structure of the underlying neural circuitry [67], which might represent visual attributes

[66], that the second order correlational structure of spontaneous and evoked activity is very similar and it changes systematically with age [64]. Thus, cell responses even in primary sensory cortices are determined by the combination of spontaneous and bottom-up, external stimulus-driven activity.

The link between spontaneous and evoked activity is further promoted by findings that after repetitive presentation of a sensory stimulus, spontaneous activity exhibits patterns of activity reminiscent to those seen during evoked activity [92]. This suggests that spontaneous activity might be altered on various time scales leading to perceptual adaptation and learning. These results led to an increasing consensus that spontaneous activity might have a functional role in perceptual processes that is related to internal states of cell assemblies in the brain, expressed via top-down effects that embody expectations, predictions and attentional processes [93] manifested in modulating functional connectivity of the network [94]. Although there have been theoretical proposals of how bottom-up and top-down signals could jointly define perceptual processes [55,95], the rigorous functional integration of spontaneous activity in such a framework has emerged only recently [53].



**Figure I.**

Characteristics of cortical spontaneous activity. **(a)** There is a significant correlation between the orientation map of the primary visual cortex of anesthetized cat (left panel), optical image patterns of spontaneous (middle panel) and visually evoked activities (right panel) (adapted with permission from [66]). **(b)** Correlational analysis of BOLD signals during resting state reveals networks of distant areas in the human cortex with coherent spontaneous fluctuations. There are large scale positive intrinsic correlations between the seed region PCC (yellow) and MPF (orange) and negative correlations between PCC and IPS (blue) (adapted with permission from [98]). **(c)** Reliably repeating spike triplets can be detected in the spontaneous firing of the rat somatosensory cortex by multielectrode recording (adapted with permission from [91]). **(d)** Spatial correlations in the developing awake ferret visual cortex of multielectrode recordings show a systematic pattern of emerging strong correlations across several millimeters of the cortical surface and very similar correlational patterns for dark spontaneous (solid line) and visually driven conditions (dotted and dashed lines for random noise patterns and natural movies, respectively) (adapted with permission from [64]).

Under a sampling-based representational account, spontaneous activity could have a natural interpretation. In a probabilistic framework, if neural activities represent samples from a distribution over external variables, this distribution must be the so-called “posterior distribution”. The posterior distribution is inferred by combining information from two sources: the sensory input, and the prior distribution describing *a priori* beliefs about the sensory environment (Figure 3b). Intuitively, in the absence of sensory stimulation, this

distribution will collapse to the prior distribution, and spontaneous activity will represent this prior (Figure 4).

This proposal linking spontaneous activity to the prior distribution has implications that can address many of the issues developed in this review. It provides an account of spontaneous activity that is consistent with one of its main features: its remarkable similarity to evoked activity [64,66,67]. A general feature of statistical models that are appropriately describing their inputs is that the prior distribution and the average posterior distribution closely match each other [68]. Thus, if evoked and spontaneous activities represent samples from the posterior and prior distributions, respectively, under an appropriate model of the environment, they are expected to be similar [53]. In addition, spontaneous activity itself, as prior expectation, should be sufficient to evoke firing in some cells without sensory input, as was observed experimentally [67].

Statistical neural networks also suggest that sampling from the prior can be more than just a byproduct of probabilistic inference: it can be computationally advantageous for the functioning of the network. In the absence of stimulation, during awake spontaneous activity, sampling from the prior can help with driving the network close to states that are probable to be valid inferences once input arrives, thus potentially shortening the reaction time of the system [69]. This “priming” effect could present an alternative account of why human subjects are able to sort images into natural/non-natural categories in a matter of ~150 ms [70], which is traditionally taken as evidence for the dominance of feed-forward processing in the visual system [71]. Finally, during off-line periods, such as sleeping, sampling from the prior could have a role in tuning synaptic weights thus contributing to the refinement of the internal model of the sensory environment as suggested by statistical neural network models [32,33].

Importantly, the proposal that spontaneous activity represents samples from the prior also provides a way to test a direct link between statistically optimal inference and learning. A match between the prior and the average posterior distribution in a statistical model is expected to develop gradually as learning proceeds [68], and this gradual match could be tracked experimentally by comparing spontaneous and evoked population activities at successive developmental stages. Such testable predictions can confirm if sampling-based representations are present in the cortex and verify the proposed link between spontaneous activity and sampling-based coding.

## Concluding remarks and future challenges

In this review, we have argued that in order to develop a unified framework that can link behavior to neural processes of both inference and learning, a key issue to resolve is the nature of neural representations of uncertainty in the cortex. We compared potential candidate neural codes that could link behavior to neural implementations in a probabilistic way by implementing computations with and learning of probability distributions of environmental features. Although explored to different extents, these coding frameworks are all promising candidates, yet each of them has shortcomings that need to be addressed in future research (Box 4). Research on PPCs needs to make viable proposals on how learning could be implemented with such representations, whereas the main challenge for sampling-based methods is to demonstrate that this scheme could work for non-trivial, dynamical cases in real time.

**Box 4****Questions for future research**

- Exact probabilistic computation in the brain is not feasible. What are the approximations that are implemented in the brain and to what extent can an approximate computation scheme still claim that it is probabilistic and optimal?
- Probabilistic learning is presently described at the neural level as a simple form of parameter learning (so-called maximum likelihood learning) at best. However, there is ample behavioral evidence for more sophisticated forms of probabilistic learning, such as model selection. These forms of learning require a representation of uncertainty about parameters, or models, not just about hidden variables. How do neural circuits represent parameter uncertainty and implement model selection?
- Highly structured neural activity in the absence of external stimulation has been observed both in the neocortex and in the hippocampus, under the headings “spontaneous activity” and “replay”, respectively. Despite the many similarities these processes show there has been little attempt to study them in a unified framework. Are the two phenomena related, is there a common function they serve?
- Can a convergence between spontaneous and evoked activities be predicted from premises that are incompatible with spontaneous activity representing samples from the prior, for example with simple correlational learning schemes?
- Can some recursive implementation of probabilistic learning link learning of low-level attributes, such as orientations, with high-level concept learning, that is, can it bridge the subsymbolic and symbolic levels of computation?
- What is the internal model according to which the brain is adapting its representation? All the probabilistic approaches have preset prior constraints that determine how inference and learning will work. Where do these constraints come from? Can they be mapped to biological quantities?

Most importantly, a tighter connection between abstract computational models and neurophysiological recordings in behaving animals is needed. For PPCs, such interactions between theoretical and empirical investigations have just begun [50]; for sampling-based methods it is still almost non-existent beyond the description of receptive fields. Present day data collection methods, such as modern imaging techniques and multi-electrode recording systems, are increasingly available to provide the necessary experimental data for evaluating the issues raised in this review. Given the complex and non-local nature of computation in probabilistic frameworks, the main theoretical challenge remains to map abstract probabilistic models to neural activity in awake behaving animals to further our understanding of cortical representations of inference and learning.

**Acknowledgments**

This work was supported by the Swartz Foundation (J.F., G.O., P.B.), by the Swiss National Science Foundation (P.B.) and the Wellcome Trust (M.L.). We thank Peter Dayan, Maneesh Sahani and Jeff Beck for useful discussions.

**Glossary**

Expected utility      the average expected reward associated with a particular decision,  $\alpha$ , when the state of the environment,  $y$ , is unknown. It can be computed

	by calculating the average of the utility function, $U(\alpha, y)$ , describing the amount of reward obtained when making decision $\alpha$ if the true state of the environment is $y$ , with regard to the <i>posterior distribution</i> , $p(y x)$ , describing the degree of belief about the state of the environment given some sensory input, $x$ : $R(\alpha) = \int U(\alpha, y) p(y x) dy$
Likelihood	the function specifying the probability $p(x y, M)$ of observing a particular stimulus $x$ for each possible state of the environment, $y$ , under a probabilistic model of the environment, $M$
Marginalization	, the process by which the distribution of a subset of variables, $y_1$ , is computed from the joint distribution of a larger set of variables, $\{y_1, y_2\}$ $p(y_1) = \int p(y_1, y_2) dy_2$ . (This could be important if, for example, different decisions rely on different subsets of the same set of variables.) Importantly, in a sampling-based representation, in which different neurons represent these different subsets of variables, simply “reading” (e.g. by a downstream brain area) the activities of only those neurons that represent $y_1$ automatically implements such a marginalization operation
Maximum <i>a posteriori</i> (or MAP) estimate	in the context of <i>probabilistic inference</i> , it is an approximation by which instead of representing the full <i>posterior</i> distribution, only a single value of $y$ is considered that has the highest probability under the <i>posterior</i> . (Formally, the full <i>posterior</i> is approximated by a Dirac-delta distribution, an infinitely narrow Gaussian, located at its maximum.) As a consequence, <i>uncertainty</i> about $y$ is no longer represented
Maximum likelihood estimate	as the <i>MAP estimate</i> , it is also an approximation, but the full <i>posterior</i> is approximated by the single value of $y$ which has the highest <i>likelihood</i>
Posterior	the probability distribution $p(y x, M)$ produced by <i>probabilistic inference</i> according to a particular probabilistic model of the environment, $M$ , giving the probability that the environment is in any of its possible states, $y$ , when stimulus $x$ is being observed
Prior	the probability distribution $p(y M)$ defining the expectation about the environment being in any of its possible states, $y$ , before any observation is available according to a probabilistic model of the environment, $M$
Probabilistic inference	the process by which the posterior is computed. It requires a probabilistic model, $M$ , of stimuli $x$ and states of the environment $y$ , containing a <i>prior</i> and a <i>likelihood</i> . It is necessary when environmental states are not directly available to the observer they can only be inferred from stimuli through inverting the relationship between $y$ and $x$ through Bayes’ rule: $p(y x, M) = p(x y, M) p(y M)/Z$ , where $Z$ is a factor independent of $y$ , ensuring that the <i>posterior</i> is a well-defined probability distribution. Note, that the <i>posterior</i> is a full probability distribution, rather than a single estimate over environmental states, $y$ . In contrast with approximate inference methods, such as <i>maximum likelihood</i> or <i>maximum a posteriori</i> that compute single best estimates of $y$ , the posterior fully represents the uncertainty about the inferred variables
Probabilistic learning	the process of finding a suitable model for <i>probabilistic inference</i> . This itself can be viewed as a problem of probabilistic inference at a higher level, where the unobserved quantity is the model, $M$ , including its

parameters and structure. Thus, the complete description of the results of probabilistic learning is a posterior distribution,  $p(M|X)$ , over possible models given all stimuli observed so far,  $X$ . Even though approximate versions, such as *maximum likelihood* or *MAP*, compute only a single best estimates of  $M$ , they still need to rely on representing uncertainty about the states of the environment,  $y$ . The effect of learning is usually a gradual change in the posterior (or estimate) as more and more stimuli are observed, reflecting the incremental nature of learning

## References

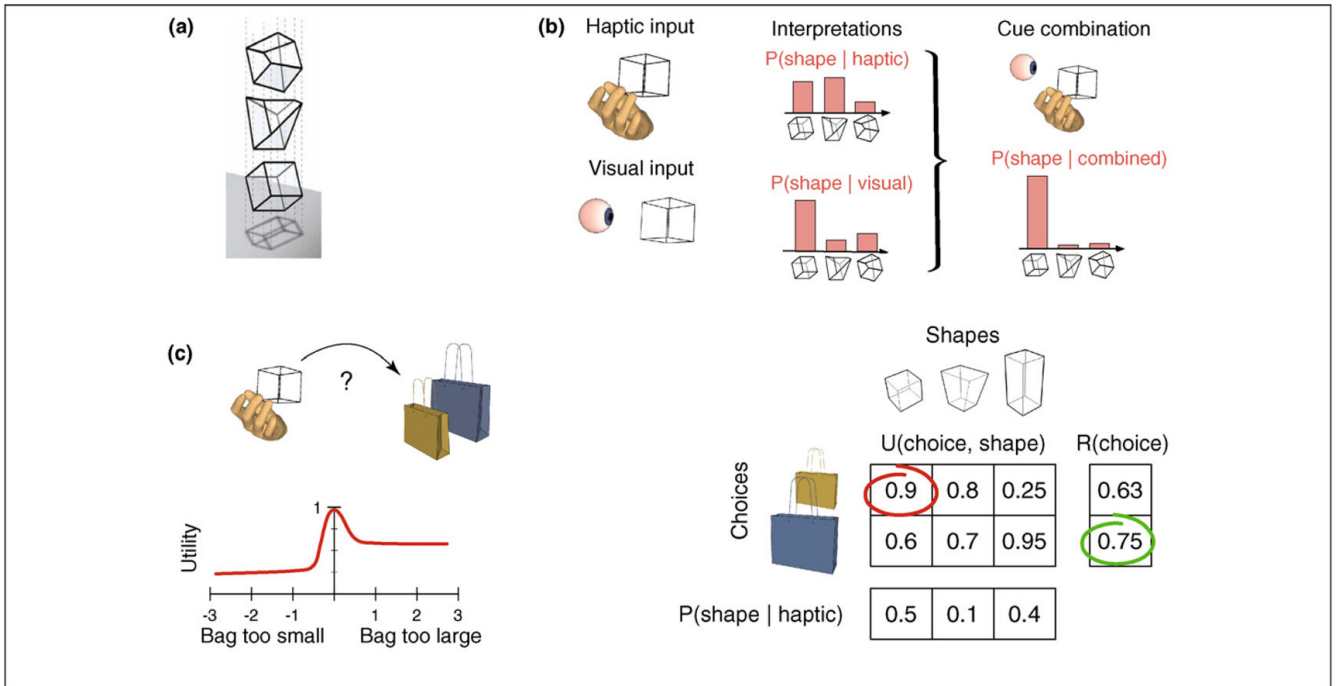
1. Barlow, HB. Possible principles underlying the transformations of sensory messages. In: Rosenblith, W., editor. *Sensory Communication*. MIT Press; 1961. p. 217-234.
2. Geisler WS, Diehl RL. Bayesian natural selection and the evolution of perceptual systems. *Philos Trans R Soc Lond B Biol Sci* 2002;357:419–448. [PubMed: 12028784]
3. Smith JD. The study of animal metacognition. *Trends Cogn Sci* 2009;13:389–396. [PubMed: 19726218]
4. Pouget A, et al. Inference and computation with population codes. *Annu Rev Neurosci* 2004;26:381–410. [PubMed: 12704222]
5. Helmholtz, HV. *Treatise on Physiological Optics*. Optical Society of America; 1925.
6. Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics*. John Wiley and Sons; 1966.
7. Cox RT. Probability, frequency and reasonable expectation. *Am J Phys* 1946;14:1–13.
8. Bernoulli, J. *Ars Conjectandi*. Thurnisiorum; 1713.
9. Atkins JE, et al. Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Res* 2001;41:449–461. [PubMed: 11166048]
10. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 2002;415:429–433. [PubMed: 11807554]
11. Kording KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature* 2004;427:244–247. [PubMed: 14724638]
12. Weiss Y, et al. Motion illusions as optimal percepts. *Nat Neurosci* 2002;5:598–604. [PubMed: 12021763]
13. Trommershäuser J, et al. Decision making, movement planning and statistical decision theory. *Trends Cogn Sci* 2008;12:291–297. [PubMed: 18614390]
14. Kording K. Decision theory: What “should” the nervous system do? *Science* 2007;318:606–610. [PubMed: 17962554]
15. Zemel R, et al. Probabilistic interpretation of population codes. *Neural Comput* 1998;10:403–430. [PubMed: 9472488]
16. Ma WJ, et al. Bayesian inference with probabilistic population codes. *Nat Neurosci* 2006;9:1432–1438. [PubMed: 17057707]
17. Beck JM, et al. Probabilistic population codes for Bayesian decision making. *Neuron* 2008;60:1142–1152. [PubMed: 19109917]
18. Jacobs RA, et al. Adaptive mixtures of local experts. *Neural Comput* 1991;3:79–87.
19. Geisler WS, et al. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res* 2001;41:711–724. [PubMed: 11248261]
20. Griffiths TL, Tenenbaum JB. Optimal predictions in everyday cognition. *Psychol Sci* 2006;17:767–773. [PubMed: 16984293]
21. Chater N, et al. Probabilistic models of cognition: conceptual foundations. *Trends Cogn Sci* 2006;10:287–291. [PubMed: 16807064]
22. Courville, AC., et al. Model uncertainty in classical conditioning. In: Thrun, S., et al., editors. *Advances in Neural Information Processing Systems 16*. MIT Press; 2004. p. 977-984.
23. Blaisdell AP, et al. Causal reasoning in rats. *Science* 2006;311:1020–1022. [PubMed: 16484500]

24. Gopnik A, et al. A theory of causal learning in children: causal maps and Bayes nets. *Psychol Rev* 2004;111:3–32. [PubMed: 14756583]
25. Kemp C, Tenenbaum JB. The discovery of structural form. *Proc Natl Acad Sci U S A* 2008;105:10687–10692. [PubMed: 18669663]
26. Orbán G, et al. Bayesian learning of visual chunks by human observers. *Proc Natl Acad Sci U S A* 2008;105:2745–2750. [PubMed: 18268353]
27. Buonomano DV, Merzenich MM. Cortical plasticity: from synapses to maps. *Annu Rev Neurosci* 1998;21:149–186. [PubMed: 9530495]
28. Gilbert CD, et al. Perceptual learning and adult cortical plasticity. *J Physiol Lond* 2009;587:2743–2751. [PubMed: 19525560]
29. Deneve S. Bayesian spiking neurons II: learning. *Neural Comput* 2008;20:118–145. [PubMed: 18045003]
30. Lengyel, M.; Dayan, P. Uncertainty, phase and oscillatory hippocampal recall. In: Scholkopf, B., et al., editors. *Advances in Neural Information Processing Systems 19*. MIT Press; 2007. p. 833-840.
31. Rumelhart, DE., et al., editors. *Parallel Distributed Processing—Explorations in the Microstructure of Cognition*. MIT Press; 1986.
32. Hinton GE, et al. The wake–sleep algorithm for unsupervised neural networks. *Science* 1995;268:1158–1161. [PubMed: 7761831]
33. Hinton, GE.; Sejnowski, TJ. Learning and relearning in Boltzmann machines. In: Rumelhart, DE.; McClelland, JL., editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press; 1986.
34. Neal, RM. *Bayesian Learning for Neural Networks*. Springer-Verlag; 1996.
35. Hinton GE. Learning multiple layers of representation. *Trends Cogn Sci* 2007;11:428–434. [PubMed: 17921042]
36. Dayan P. Recurrent sampling models for the Helmholtz machine. *Neural Comput* 1999;11:653–677. [PubMed: 10085425]
37. Dayan P, Hinton GE. Varieties of Helmholtz machine. *Neural Netw* 1996;9:1385–1403. [PubMed: 12662541]
38. Hinton GE, Ghahramani Z. Generative models for discovering sparse distributed representations. *Philos Trans R Soc Lond B Biol Sci* 1997;352:1177–1190. [PubMed: 9304685]
39. Lucke J, Sahani M. Maximal causes for non-linear component extraction. *J Mach Learn Res* 2008;9:1227–1267.
40. Bell AJ, Sejnowski TJ. The “independent components” of natural scenes are edge filters. *Vision Res* 1997;37:3327–3338. [PubMed: 9425547]
41. Berkes P, Wiskott L. Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis* 2005;5:579–602. [PubMed: 16097870]
42. Karklin Y, Lewicki MS. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 2009;457:83–85. [PubMed: 19020501]
43. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 1996;381:607–609. [PubMed: 8637596]
44. Rao RPN, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 1999;2:79–87. [PubMed: 10195184]
45. Roweis S, Ghahramani Z. A unifying review of linear Gaussian models. *Neural Comput* 1999;11:305–345. [PubMed: 9950734]
46. Turner R, Sahani M. A maximum-likelihood interpretation for slow feature analysis. *Neural Comput* 2007;19:1022–1038. [PubMed: 17348772]
47. Schwartz O, Simoncelli EP. Natural signal statistics and sensory gain control. *Nat Neurosci* 2001;4:819–825. [PubMed: 11477428]
48. Olshausen BA, Field DJ. How close are we to understanding V1? *Neural Comput* 2005;17:1665–1699. [PubMed: 15969914]
49. Lopez-Aranda MF, et al. Role of layer 6 of V2 visual cortex in object-recognition memory. *Science* 2009;325:87–89. [PubMed: 19574389]



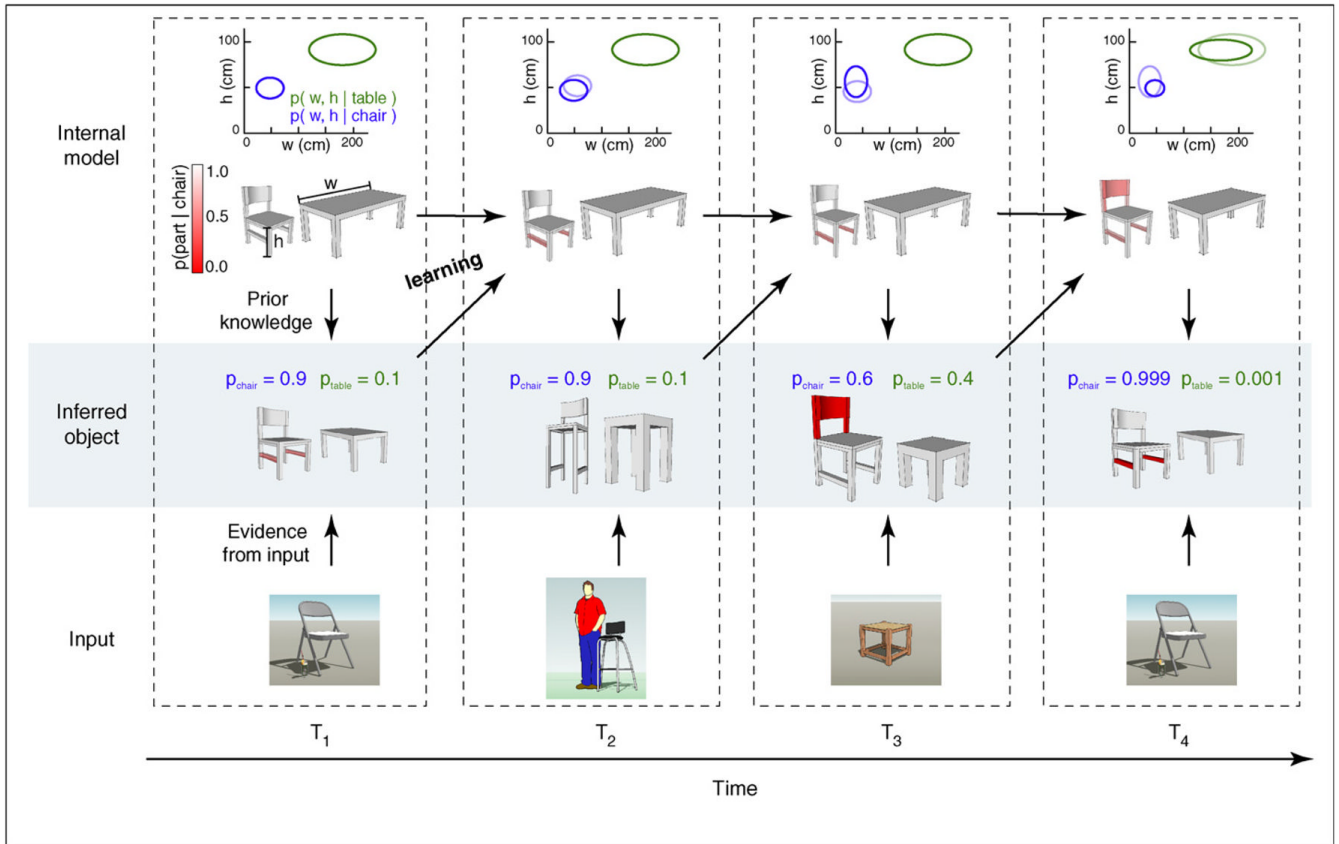
50. Yang T, Shadlen MN. Probabilistic reasoning by neurons. *Nature* 2007;447:1075–1080. [PubMed: 17546027]
51. Anderson, CH.; Van Essen, DC. Neurobiological computational systems. In: Zurenda, JM., et al., editors. *Computational Intelligence Imitating Life*. IEEE Press; 1994. p. 213-222.
52. Ma WJ, et al. Spiking networks for Bayesian inference and choice. *Curr Opin Neurobiol* 2008;18:217–222. [PubMed: 18678253]
53. Berkes, P., et al. Matching spontaneous and evoked activity in V1: a hallmark of probabilistic inference. *Frontiers in Systems Neuroscience*. Conference Abstract: Computational and systems neuroscience; 2009.
54. Hoyer, PO.; Hyvarinen, A. Interpreting neural response variability as Monte Carlo sampling of the posterior. In: Becker, S., et al., editors. *Advances in Neural Information Processing Systems 15*. MIT Press; 2003. p. 277-284.
55. Lee TS, Mumford D. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am (A)* 2003;20:1434–1448.
56. Zohary E, et al. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 1994;370:140–143. [PubMed: 8022482]
57. Leopold DA, Logothetis NK. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 1996;379:549–553. [PubMed: 8596635]
58. Sundareswara R, Schrater PR. Perceptual multistability predicted by search model for Bayesian decisions. *J Vis* 2008;8:12.1–12.19. [PubMed: 18842083]
59. Daw, N.; Courville, A. The rat as particle filter. In: Platt, J., et al., editors. *Advances in Neural Information Processing Systems 20*. MIT Press; 2008. p. 369-376.
60. Levy, RP.; Reali, F.; Griffiths, TL. Modeling the effects of memory on human online sentence processing with particle filters. In: Koller, P., et al., editors. *Advances in Neural Information Processing Systems 21*. MIT Press; 2009. p. 937-944.
61. Vul E, Pashler H. Measuring the crowd within—probabilistic representations within individuals. *Psychol Sci* 2008;19:645–647. [PubMed: 18727777]
62. Shadlen MN, Newsome WT. Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 1994;4:569–579. [PubMed: 7812147]
63. Anderson JS, et al. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science* 2000;290:1968–1972. [PubMed: 11110664]
64. Fiser J, et al. Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature* 2004;431:573–578. [PubMed: 15457262]
65. Ringach D. Spontaneous and driven cortical activity: implications for computation. *Curr Opin Neurobiol* 2009;19:439–444. [PubMed: 19647992]
66. Kenet T, et al. Spontaneously emerging cortical representations of visual attributes. *Nature* 2003;425:954–956. [PubMed: 14586468]
67. Tsodyks M, et al. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 1999;286:1943–1946. [PubMed: 10583955]
68. Dayan, P.; Abbott, LF. *Theoretical Neuroscience*. MIT Press; 2001.
69. Neal RM. Annealed importance sampling. *Stat Comput* 2001;11:125–139.
70. Thorpe S, et al. Speed of processing in the human visual system. *Nature* 1996;381:520–522. [PubMed: 8632824]
71. Serre T, et al. A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U S A* 2007;104:6424–6429. [PubMed: 17404214]
72. Saffran JR, et al. Statistical learning by 8-month-old infants. *Science* 1996;274:1926–1928. [PubMed: 8943209]
73. Seitz AR, et al. Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception* 2007;36:1445–1453. [PubMed: 18265827]
74. Pena M, et al. Signal-driven computations in speech processing. *Science* 2002;298:604–607. [PubMed: 12202684]
75. Turk-Browne NB, et al. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci* 2009;21:1934–1945. [PubMed: 18823241]

76. Fiser J, Aslin RN. Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 2001;12:499–504. [PubMed: 11760138]
77. Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. *Proc Natl Acad Sci U S A* 2002;99:15822–15826. [PubMed: 12429858]
78. Kirkham NZ, et al. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 2002;83:B35–B42. [PubMed: 11869728]
79. Fiser J, Aslin RN. Encoding multielement scenes: statistical learning of visual feature hierarchies. *J Exp Psychol Gen* 2005;134:521–537. [PubMed: 16316289]
80. Beck JM, Pouget A. Exact inferences in a neural implementation of a hidden Markov model. *Neural Comput* 2007;19:1344–1361. [PubMed: 17381269]
81. Deneve S. Bayesian spiking neurons I: inference. *Neural Comput* 2008;20:91–117. [PubMed: 18045002]
82. Huys QJM, et al. Fast population coding. *Neural Comput* 2007;19:404–441. [PubMed: 17206870]
83. Creutzfeldt OD, et al. Relations between EEG phenomena and potentials of single cortical cells: II. Spontaneous and convulsoid activity. *Electroencephalogr Clin Neurophysiol* 1966;20:19–37. [PubMed: 4161316]
84. Tolhurst DJ, et al. The statistical reliability of signals in single neurons in cat and monkey visual-cortex. *Vision Res* 1983;23:775–785. [PubMed: 6623937]
85. Wu JY, et al. Propagating waves of activity in the neocortex: what they are, what they do. *Neuroscientist* 2008;14:487–502. [PubMed: 18997124]
86. Katz LC, Shatz CJ. Synaptic activity and the construction of cortical circuits. *Science* 1996;274:1133–1138. [PubMed: 8895456]
87. Arieli A, et al. Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *J Neurophysiol* 1995;73:2072–2093. [PubMed: 7623099]
88. Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci* 2007;8:700–711. [PubMed: 17704812]
89. Attwell D, Laughlin SB. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab* 2001;21:1133–1145. [PubMed: 11598490]
90. Ikegaya Y, et al. Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 2004;304:559–564. [PubMed: 15105494]
91. Luczak A, et al. Sequential structure of neocortical spontaneous activity in vivo. *Proc Natl Acad Sci U S A* 2007;104:347–352. [PubMed: 17185420]
92. Han F, et al. Reverberation of recent visual experience in spontaneous cortical waves. *Neuron* 2008;60:321–327. [PubMed: 18957223]
93. Gilbert CD, Sigman M. Brain states: top-down influences in sensory processing. *Neuron* 2007;54:677–696. [PubMed: 17553419]
94. Nauhaus I, et al. Stimulus contrast modulates functional connectivity in visual cortex. *Nat Neurosci* 2009;12:70–76. [PubMed: 19029885]
95. Yuille A, Kersten D. Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 2006;10:301–308. [PubMed: 16784882]
96. Ernst MO, Bulthoff HH. Merging the senses into a robust percept. *Trends Cogn Sci* 2004;8:162–169. [PubMed: 15050512]
97. Kording KP, Wolpert DM. Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 2006;10:319–326. [PubMed: 16807063]
98. Fox MD, et al. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A* 2005;102:9673–9678. [PubMed: 15976020]
99. Berkes P, et al. A structured model of video reproduces primary visual cortical organisation. *PLoS Comput Biol* 2009;5:e1000495. [PubMed: 19730679]
100. Rossi AF, et al. The representation of brightness in primary visual cortex. *Science* 1996;273:1104–1107. [PubMed: 8688096]
101. Adelson EH. Perceptual organization and the judgment of brightness. *Science* 1993;262:2042–2044. [PubMed: 8266102]



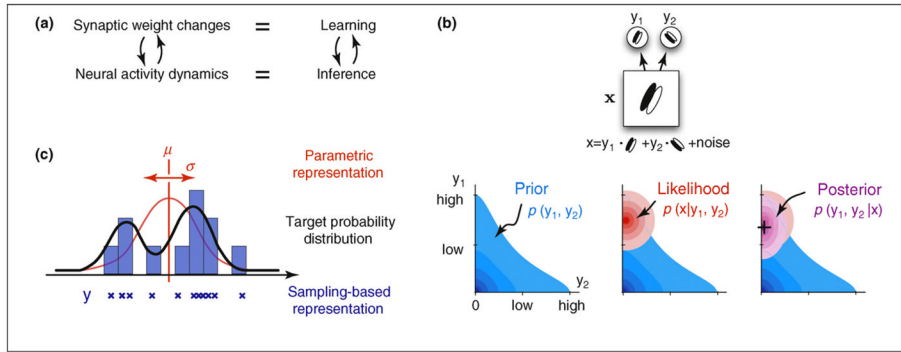
**Figure 1.**

Representation of uncertainty and its benefits. **(a)** Sensory information is inherently ambiguous. Given a two-dimensional projection on a surface (e.g. a retina), it is impossible to determine which of the three different three-dimensional wire frame objects above cast the image (adapted with permission from [96]). **(b)** Cue integration. Independent visual and haptic measurements (left) support to different degrees the three possible interpretations of object identity (middle). Integrating these sources of information according to their respective uncertainties provides an optimal probabilistic estimate of the correct object (right). **(c)** Decision-making. When the task is to choose the bag with the right size for storing an object, uncertain haptic information needs to be utilized probabilistically for optimal choice (top left). In the example shown, the utility function expresses the degree to which a combination of object and bag size is preferable: for example, if the bag is too small, the object will not fit in, if it is too large, we are wasting valuable bag space (bottom left, top right). In this case, rather than inferring the most probable object based on haptic cues and then choosing the bag optimal for that object (in the example, the small bag for the cube), the probability of each possible object needs to be weighted by its utility and the combination with the highest expected utility (R) has to be selected (in the example, the large bag has the highest expected utility). Evidence shows that human performance in cue combination and decision-making tasks is close to optimal [10,97].

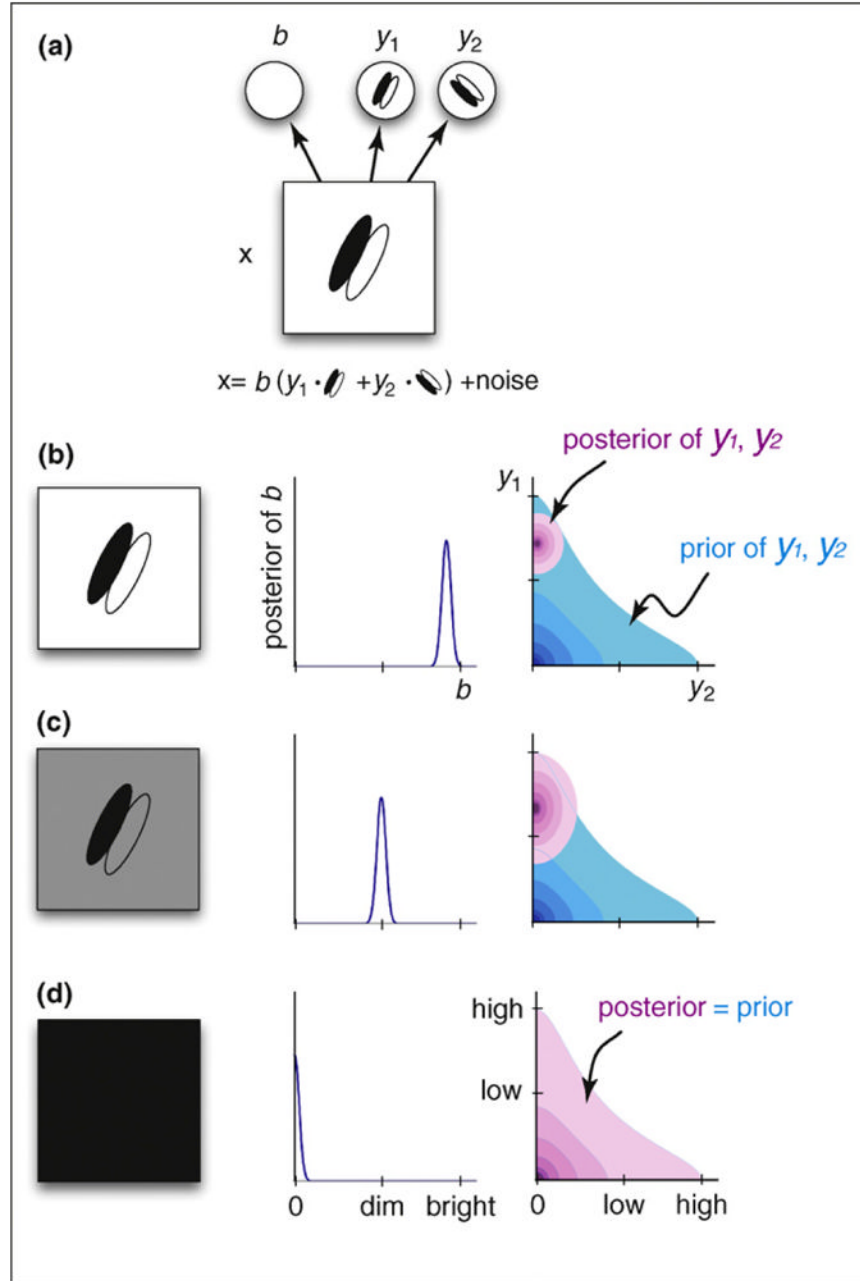


**Figure 2.**

The link between probabilistic inference and learning. (Top row) Developing internal models of chairs and tables. The plot shows the distribution of parameters (two-dimensional Gaussians, represented by ellipses) and object shapes for the two categories. (Middle row) Inferences about the currently viewed object based on the input and the internal model. (Bottom row) Actual sensory input. Red color code represents the probability of a particular object part being present (see color scale on top left). T1 –T4, four successive illustrative iterations of the inference – learning cycle. (T1) The interpretation of a natural scene requires combining information from the sensory input (bottom) and the internal model (top). Based on the internal models of chairs and tables, the input is interpreted with high probability ( $p = 0.9$ ) as a chair with a typical size but missing crossbars (middle). (T2) The internal model of the world is updated based on the cumulative experience of previous inferences (top). The chair in T1, being a typical example of a chair, requires minimal adjustments to the internal model. Experience with more unusual instances, as for example the high chair in T2, provokes more substantial changes (T3, top). (T3) The representation of uncertainty allows to update the internal model taking into account all possible interpretations of the input. In T3, the stimulus is ambiguous as it could be interpreted as a stool, or a square table. The internal model needs to be updated by taking into account the relative probability of the two interpretations: that there exist tables with a more square shape or that some chairs miss the upper part. Since both probabilities are relatively high, both internal models will be modified substantially during learning (see the change of both ellipses). (T4) After learning, the same input as in T1 elicits different responses owing to changes in the internal model. In T4, the input is interpreted as a chair with significantly higher confidence, as experience has shown that chairs often lack the bottom crossbars.



**Figure 3.** Neural substrates of probabilistic inference and learning. **(a)** Functional mapping of learning and inference onto neural substrates in the cortex. **(b)** Probabilistic inference for natural images. (Top) A toy model of the early visual system (based on Ref. [43]). The internal model of the environment assumes that visual stimuli,  $x$ , are generated by the noisy linear superposition of two oriented features with activation levels,  $y_1$  and  $y_2$ . The task of the visual system is to infer the activation levels,  $y_1$  and  $y_2$ , of these features from seeing only their superposition,  $x$ . (Bottom left) The *prior distribution* over the activation of these features,  $y_1$  and  $y_2$ , captures prior knowledge about how much they are typically (co-)activated in images experienced before. In this example,  $y_1$  and  $y_2$  are expected to be independent and sparse, which means that each feature appears rarely in visual scenes and independently of the other feature. (Bottom middle) The *likelihood* function represents the way the visual features are assumed to combine to form the visual input under our model of the environment. It is higher for feature combinations that are more likely to underlie the image we are seeing according to the equation on the top. (Bottom right) The goal of the visual system is to infer the *posterior distribution* over  $y_1$  and  $y_2$ . By Bayes’ theorem, the posterior optimally combines the expectations from the prior with the evidence from the likelihood. *Maximum a posteriori* (MAP) estimate, used by some models [40,43,47], denoted by a + in the figure neglects uncertainty by using only the maximum value instead of the full distribution. **(c)** Simple demonstration of two probabilistic representational schemes. (Black curve) The probability distribution of variable  $y$  to be represented. (Red curve) Assumed distribution by the parametric representation. Only the two parameters of the distribution, the mean  $\mu$  and variance  $\sigma$  are represented. (Blue “x”-s and bars) Samples and the histogram implied by the sampling-based representation.



**Figure 4.**

Relating spontaneous activity in darkness to sampling from the prior, based on the encoding of brightness in the primary visual cortex. (a) A statistically more efficient toy model of the early visual system [47,99] (Figure 3b). An additional feature variable,  $b$ , has a multiplicative effect on other features, effectively corresponding to the overall luminance. Explaining away this information removes redundant correlations thus improving statistical efficiency. (b–c) Probabilistic inference in such a model results in a luminance-invariant behavior of the other features, as observed neurally [100] as well as perceptually [101]: when the same image is presented at different global luminance levels (left), this difference is captured by the posterior distribution of the “brightness” variable,  $b$  (center), whereas the posterior for other features, such as  $y_1$  and  $y_2$ , remains relatively unaffected (right). (d) In the limit of total darkness (left),

the same luminance-invariant mechanism results in the posterior over  $y_1$  and  $y_2$  collapsing to the prior (right). In this case, the inferred brightness,  $b$ , is zero (center) and as  $b$  explains all of the image content, there is no constraint left for the other feature variables,  $y_1$  and  $y_2$  (the identity in  $\mathbf{a}$  becomes  $0 = 0 (y_1 \cdot w_1 + y_2 \cdot w_2)$ , which is fulfilled for every value of  $y_1$  and  $y_2$ ).

**Table I**

Comparing characteristics of the two main modeling approaches to probabilistic neural representations

	<b>PPCs</b>	<b>Sampling-based</b>
Neurons correspond to	Parameters	Variables
Network dynamics required (beyond the first layer)	Deterministic	Stochastic (self-consistent)
Representable distributions	Must correspond to a particular parametric form	Can be arbitrary
Critical factor in accuracy of encoding a distribution	Number of neurons	Time allowed for sampling
Instantaneous representation of uncertainty	Complete, the whole distribution is represented at any time	Partial, a sequence of samples is required
Number of neurons needed for representing multimodal distributions	Scales exponentially with the number of dimensions	Scales linearly with the number of dimensions
Implementation of learning	Unknown	Well-suited