

The final version of this article will appear in *Neural Computation*, Vol. 15, Issue 10, published by The MIT Press.

Doubly Distributional Population Codes: Simultaneous Representation of Uncertainty and Multiplicity

Maneesh Sahani* **Peter Dayan**
Gatsby Computational Neuroscience Unit
University College London
17 Queen Square,
London, WC1N 3AR, UK.

Abstract

Perceptual inference fundamentally involves uncertainty, arising from noise in sensation and the ill-posed nature of many perceptual problems. Accurate perception requires that this uncertainty be correctly represented, manipulated, and learned about. The choices made by subjects in various psychophysical experiments suggest that they do indeed take such uncertainty into account when making perceptual inferences, posing the question as to how uncertainty is represented in the activities of neuronal populations. Most theoretical investigations of population coding have ignored this issue altogether; the few existing proposals that address it, do so in such a way that it is fatally conflated with another facet of perceptual problems that also needs correct handling, namely multiplicity (that is, the simultaneous presence of multiple distinct stimuli). We present and validate a more powerful proposal for the way that population activity may encode uncertainty, both distinctly from, and simultaneously with, multiplicity.

*Present address: Keck Center for Integrative Neuroscience, Box 0732, University of California, 513 Parnassus Ave, San Francisco, CA 94143-0732

1 Introduction

One of the few widely accepted facts about vertebrate neural coding is that representation of both input and output involves population codes (*e.g.*, Pouget *et al.* 2000). Single neurons in cortical areas such as V1 or M1 are broadly tuned, that is, their firing rates vary in response to a fairly broad spectrum of different stimuli, such as different directions of motion of a single bar of light (Hubel and Wiesel 1968) or different trajectories of an upcoming arm movement (Georgopoulos *et al.* 1986). Thus information about a stimulus is available in the activity of many neurons. This coding strategy is evidently robust to noise and cell death; its other computational characteristics and advantages have been the subject of much investigation (Hinton 1984; Snippe and Koenderink 1992; Zhang and Sejnowski 1999; Eurich and Wilke 2000).

Although the idea that information is represented in the activities of whole populations of neurons is now commonplace, there is an active debate about exactly what information is represented and how. Early work concentrated on the representation of a single value s of a physical stimulus feature, such as the direction of motion of a moving spot, orientation of a bar, or the location of the source of a sound. In this case, the way that neuron i encodes s can be characterized by its *tuning curve* $f_i(s)$, which gives the mean firing rate for each value of the feature. The actual firing rate r_i of the cell is randomly distributed about this mean (possibly with correlations between different cells). We use the symbol \sim to indicate this relationship, and write

$$r_i \sim f_i(s). \quad (1)$$

Given this model of the way that s is *encoded* in $\mathbf{r} = \{r_i\}$, it is straightforward to understand how s can be *decoded* from the activities, how much information is contained in \mathbf{r} about s , and even what tuning properties of cells best serve various goals (Paradiso 1988; Baldi and Heiligenberg 1988; Snippe and Koenderink 1992; Seung and Sompolinsky 1993; Zhang and Sejnowski 1999; Pouget *et al.* 1999; Eurich and Wilke 2000; Dayan and Abbott 2001; Wilke and Eurich 2002)

The experiments that inspired the encoding model of equation 1 involved relatively simple scenarios in which a single-valued stimulus was presented (or, for the motor system, a single movement elicited). However, in a more realistic setting, two complications arise that increase the demands on the representational capacity of population codes. The first, *multiplicity*, occurs when multiple values of s need to be represented simultaneously. This may happen, for example, when overlapping transparent surfaces are moving in different directions (usually simulated in experiments by overlaid fields of moving dots), or when different sounds are produced by multiple spatially separated sources. Common experience, as well as a number of psychophysical experiments (*e.g.*, Treue *et al.* 2000; Feng and Ratnam 2000), shows that observers are able to perceive stimuli of this sort correctly, and so neural population codes must be capable of representing them.

The second complication is *uncertainty*, which arises in two rather different ways: noise and the ill-posed nature of many perceptual inference problems. Even in cases in which s is the particular aspect of the stimulus that is directly or nearly directly represented by neural activities in a sense organ (such as the intensity of sound around a particular frequency), these activities will be corrupted by detector noise, and their cortical reports will be corrupted by spike- and synapse-associated transmission noise. Worse than this, though, is that in most cases, s is a quantity that has to be computed from sensory neuron activities, and a complete description of the results of these computations almost always involves uncertainty about s . A classic example of this latter form of uncertainty is the aperture problem in motion perception: if s is the direction of visual motion at a particular point in space, then local measurements give rise to uncertainty depending on the spatial structure of the scene (Horn and Schunck 1981).

We are not generally aware of uncertainty in the final result of perception; however, many experiments, particularly those involving multiple sources of differentially reliable information, have shown that subjects do indeed take into account uncertainty in intermediate quantities while making certain perceptual inferences. Examples include integrating local motion information, uncertain due to the aperture problem mentioned above, to infer the overall direction of a moving object (Weiss and Adelson 1998); integrating visual and haptic information to judge the height of a ridge (Ernst and Banks 2002); and integrating visual cues

to judge the depth of a virtual cylinder (Atkins *et al.* 2001) or the orientation of a skew-symmetric surface (Saunders and Knill 2001). Additionally, computational models designed to account for other psychophysical phenomena have often posited the representation and manipulation of uncertainty in inferred quantities, for example in a model of feature-binding used to account for the phenomenon of illusory conjunctions (Ashby *et al.* 1996).

Both multiplicity and uncertainty may be viewed as types of distribution of the stimulus variable s . Anderson (1990) and Zemel *et al.* (1998) pointed out that the activities \mathbf{r} of the standard encoding model of equation 1 are fundamentally based on only a single value of s , and, as such, are unable to encode such distributional objects. Of course, *decoding* the stimulus feature from \mathbf{r} does lead to uncertainty about s , if \mathbf{r} is noisy, but Zemel *et al.* (1998) showed that this decoding uncertainty is unsuitable as a method of encoding uncertainty. These authors suggested different versions of a more powerful alternative, called a distributional population code (DPC). Such a code uses a different encoding (and therefore decoding) model, viewing the activity of the population as representing a whole *function*, $m(s)$, over s , rather than only a single value of s . Importantly, if the function is formally trivial, *i.e.* $m(s)$ is a delta function, being non-zero at only a single value of s , this is exactly the same as encoding just that single value. Thus, the new scheme can be viewed as an exact generalization of previous models of population coding. However, the function-based encoding of the DPC is far more powerful, with $m(s)$ being able to represent *either* multiplicity *or* (if normalized) uncertainty in s . That DPCs cannot concurrently represent or distinguish both, is the main motivation for our new proposal.

In a DPC, what amounts to the tuning curve of each neuron is separated into two components: a linear response function, $f_i(\cdot)$ (which might, unlike a tuning curve, be negative for some input values) that multiplies the function $m(s)$ to be encoded, and a static non-linearity $\sigma_i(\cdot)$. The firing rates are assumed to be given by

$$r_i \sim \sigma_i \left(\int ds f_i(s) m(s) \right). \quad (2)$$

When $m(s)$ is a delta function this reduces to the simple population code of equation 1, with the tuning curve given by $\sigma_i(f_i(\cdot))$. When $m(s)$ is more elaborate, the activities \mathbf{r} of equation 2 provide sufficient information to decode an approximation to $m(s)$, in the same way that it is possible to decode an approximation to s from the activities \mathbf{r} of the encoding model of equation 1. Of course, since $m(s)$ is a more complicated object than s alone, decoding too is both conceptually and computationally somewhat more complicated than for the standard encoding model.

The DPC framework is consistent with experimental measurements of neuronal responses to stimuli with multiplicity, in the form of area MT activity evoked by multiple moving random-dot fields (Treue *et al.* 2000; Zemel and Dayan 2000). The situation with regard to stimuli that would be likely to invoke uncertainty is less clear. As mentioned above, the DPC approach suffers from the crucial ambiguity of employing exactly the same mechanism to represent both multiplicity and uncertainty, and is therefore unable to distinguish between the two different types of distributional object. This makes it unsuitable for situations in which either multiplicity or uncertainty (or both together) may be relevant to the same inferred stimulus dimension, particularly when the distributions of multiplicity and uncertainty may take similar functional forms. Two examples of such situations arise in the encoding of the azimuthal location of a sound source, and the direction of motion of a moving field of dots.

Multiplicity in sound localization results simply from the presence of multiple sources of sound (Takahashi and Keller 1994). Uncertainty in sound localization arises in two different ways. One is noise in sound detection, which is very different from multiplicity. The second is potentially confusable with multiplicity, and arises computationally. In some frequency ranges, interaural phase disparity is an important cue for localization in azimuth. For narrow-band stimuli, the interaural phase disparity only specifies the difference in path length from the source to the two ears up to an integral number of wavelengths. This leads to a multimodal distribution of possible sound locations, with a mode for each position at which the distance from the two ears is consistent with the interaural phase disparity. This uncertainty resembles in functional form a distribution that might describe multiple simultaneous sources of sound. This problem posed by computational uncertainty is particularly acute for animals such as the barn owl which use interaural phase

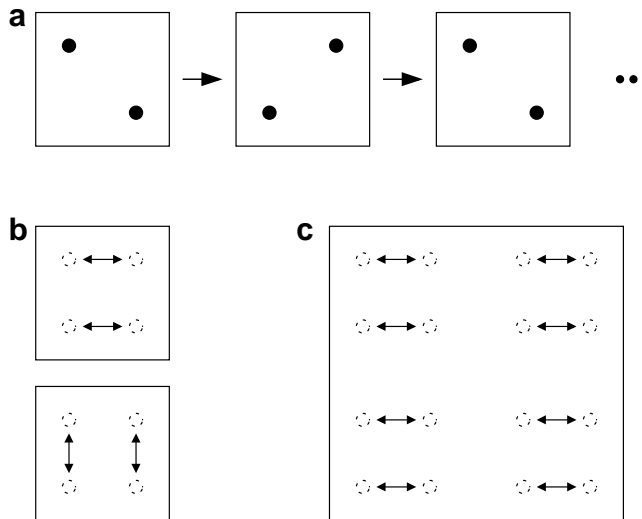


Figure 1: Apparent motion and the correspondence problem. **a**: Frames of a stimulus demonstrating bimodal motion uncertainty. **b**: Interpretations consistent with the stimulus in **a**. **c**: in field of such stimuli, only one direction of motion is perceived at a time.

disparity to localize relatively high-frequency sounds (Saber *et al.* 1999).

In visual motion detection, multiplicity can arise from transparent surfaces sliding over each other, or (at least in experimental stimuli) from interspersed random fields of dots that move in different directions (Treue *et al.* 2000). A form of uncertainty that is similar to this arises in the perception of apparent motion, due to what is called the correspondence problem. Consider a multiple-frame stimulus in which two dots appear at opposite corners of an invisible square, subsequently jump to the other pair of corners, and continue to jump to-and-fro (figure 1). The association of points in the two frames is ambiguous. As in many other cases of perceptual uncertainty, perception of such a stimulus is bistable, with only one or the other of the two percepts experienced at any one time. However, Ramachandran and Anstis (1985) have shown that in a field of such oscillating dots, all the component dots are perceived to move in the same direction. This suggests that the uncertain motion is integrated across the scene, a computation that most likely requires the local uncertainty to be represented. A more familiar form of motion uncertainty arises locally due to the aperture problem, in which the motion of a bar in a direction parallel to its orientation cannot be determined from the moving image falling within a restricted receptive field. Local uncertainty, due to the limited size of neuronal receptive fields, is resolved through integration of local signals using form cues. Weiss and Adelson (1998) have argued that such motion integration is best modelled, over a wide range of stimulus regimes, in Bayesian terms; such models require that complete information about the local uncertainty be represented, than, say, just the locally most probable direction.

In this paper, we extend the distributional population coding framework to distinguish between distributions coming from multiplicity and uncertainty. Indeed, our method can represent both together, in cases where inference leads to uncertainty about a potentially multi-valued quantity. We achieve this capability by introducing a *doubly distributional population code* (DDPC) which is based on encoding uncertainty about the functions $m(s)$ of s , rather than encoding uncertainty about s itself in the shape of just a single function $m(s)$. Formally, the relationship between distributional and doubly distributional population codes is the same as that between regular population codes and distributional population codes.

2 Results

2.1 Encoding

The standard model for population codes suggests how the noisy activities r_i of a population of neurons can represent a single value of some external stimulus variable s , like a direction of motion in the input. Distributional population codes go one stage further, and show how to encode a distribution function $m(s)$ over s , providing a substrate for the encoding and decoding of functions. Doubly distributional population codes go yet one step more, suggesting that the activities r_i actually encode a probability distribution $p[m]$ over *functions* $m(s)$ of s . Here the function $m(s)$ captures potential multiplicity; the distribution $p[m]$ captures potential uncertainty. As an example, in the case of motion transparency, the function might indicate the strength in the input of the motion in each possible direction; the distribution over such functions then conveys the uncertainty about the sets of strengths. The failing of distributional population codes is to treat these two very different distributional objects as though they were the same.

Multiplicity, $m(s)$, and uncertainty, $p[m]$, arise from quite different computations, and this difference is vital to the proposed encoding scheme. For simple sensory features, multiplicity results directly from the multiple activations of sensory receptors. In the experiment of Treue *et al.* for example, dots moving in many different directions were shown, leading to the simultaneous activation of local-motion sensitive cells with many different preferred directions early in the visual pathway. Thus, the encoding scheme of distributional population coding applies naturally; as the local-motion detectors' outputs are pooled, the inputs to each higher-level neuron combine (at least approximately) linearly, and this weighted sum (possibly represented by a somatic potential) is converted to a firing rate through a nonlinear transform. In this way, a perfectly certain function $m(s)$ would be coded by neuron i , with linear filter function $f_i(s)$ and non-linear transfer function $\sigma_i(\cdot)$ in its noisy activity, whose mean, as in equation 2, is:

$$\phi_i[m] = \sigma_i \left(\int ds f_i(s) m(s) \right). \quad (3)$$

By contrast, uncertainty about $m(s)$ arises in a different manner. Take the issue of ill-posedness, as in the aperture problem. The same inputs (motion energy in a small receptive field) can arise for different possible underlying causes (overall directions of motion). For further motion processing to be accurate, the whole ensemble of possible causes, embodied in the distribution $p[m]$, must be represented by the population code. In general, of course, the particular form of $p[m]$ depends on the particular ambiguities of the stimulus. These can be learned over multiple encounters with similar stimuli if, for instance, disambiguating information is available from other sensory inputs (other parts of the image). We argue in the Discussion that the result of such learning will be for the firing rate of the neurons to tend to their averages under the distribution of uncertainty. More precisely, the mean activity of neuron i comes from the expected value of the DPC activity $\phi_i[m]$ for function $m(s)$, averaged over the inferential distribution $p[m]$ over such functions:

$$r_i(p[m]) \sim \langle \phi_i[m] \rangle_{p[m]} = \left\langle \sigma_i \left(\int ds f_i(s) m(s) \right) \right\rangle_{p[m]}. \quad (4)$$

The actual activity of neuron i for a given input will be corrupted by noise, just as in equation 1. We refer to this representation of the distribution $p[m]$ in neuronal firing rates as a doubly distributional population code (DDPC). Just like equation 1, equation 4 is a phenomenological model of the way that stimulus information is encoded in noisy activities \mathbf{r} . It can therefore be validated only by experimental evidence. When there is no uncertainty about the function $m(s)$, that is, $p[m]$ is a δ -distribution about some particular function, the DDPC of equation 4 reduces to the DPC given by equation 2. And, as for the DPC, when only one unambiguous value is to be represented, it further reduces to the simple population code of equation 1. In these cases, then, it is consistent with previous experimental data, including those of Treue *et al.* that show the coding of multiplicity. To our knowledge, there has yet to be a direct test of the nature of the population codes that arise in the face of uncertainty about inferred quantities.

Pending such a test, and as is standard for conventional population codes, we construct an explicit decoding scheme to show that the encoding scheme in equation 4 is sufficient to represent both types of distributional object and to distinguish between them. In particular, we show that it is possible to represent distinctly both a perfectly known, multivalued stimulus (roughly, simple $p[m]$, complex $m(s)$, as in Treue *et al*), and multimodal uncertainty about a single value (simple $m(s)$, complex $p[m]$, as in ambiguous apparent motion). To achieve such a rich representation with the same elements as used by the DPC encoding scheme (namely the tuning functions f_i and the transfer functions σ_i), we must consider a larger population of cells, associating a number of different transfer functions with the same underlying tuning functions.

Figure 2 shows the mean firing rates for a population of neurons encoding each of these two distributional objects in the context of an underlying angular stimulus variable s (which might correspond to either the azimuth of a sound source or the direction of a moving object). The linear response functions of the neurons are all Gaussian-shaped curves of uniform height and width, with centers (*i.e.*, preferred directions for the neurons) that fall at one of 51 regularly spaced points along the dimension s . A group of 10 neurons shares each one of these preferred directions, but the neurons in each such group have threshold-linear transfer functions with different, regularly spaced, thresholds; the slope of the linear part of the transfer function is the same for all the neurons. The population thus consists of a total of 510 cells. Figure 2a shows the tuning curves $\sigma_i(f_i(s))$ for the group of 10 cells that have a preferred direction of 0° . If the abscissa were relabelled to read “preferred direction,” then these same curves would serve to show the response of the entire population to a single, completely certain, stimulus with direction $s = 0^\circ$; each of the 10 curves would now represent the firing rates of the 51 neurons with different preferred directions but identical thresholds.

Figure 2d–e show the DDPCs for a multivalued and an uncertain stimulus, respectively. In figure 2d, the DDPC encodes a single multiplicity function $m(s)$, given by the sum of two delta functions located at $+45^\circ$ and -45° (shown in figure 2b). In figure 2e two equiprobable, single-valued, multiplicity functions are encoded, one with a stimulus only at $+45^\circ$ and the other with a stimulus only at -45° (figure 2c). The two possible functions in the uncertain stimulus are scaled so that the expected value under the uncertainty (figure 2c) is equal to that of the (certain) multivalued function (figure 2b).

Despite the equality of expected values of the two encoded distributions, the population firing rates in the two cases differ. We now show that this difference is sufficient to identify each of the two encoded distributions from noisy versions of these rates.

2.2 Decoding

Given a pattern of uncertainty encoded in a population, the nervous system need only perform further inference or carry out other computations based on this uncertainty. The activity of the population need not be “decoded” (that is, re-represented in a more explicit form) for these computations to be carried out. However, despite its biological irrelevance, an explicit decoding algorithm can be used to offer a manifest demonstration of the representational power of the code.

It is important to assess at the outset what we might hope to be able to determine about the distribution $p[m]$ from the neural activities \mathbf{r} that encode it. We are limited by two factors. First, $p[m]$ is an infinite-dimensional object that is represented in the (discrete) activity of only a finite number of cells. This poses a standard regularization problem, which we will solve in the standard way of imposing a prior distribution $P(p[m])$ over the encoded uncertainty. Second, since the firing of neurons, particularly in the neocortex, appears to be noisy, some information about $p[m]$ available in the mean rates of equation 4 will be lost in the stochasticity. We handle this by the maximum *a posteriori* procedure of attempting to identify the distribution $q[m]$ which is most likely to have been encoded, given the actual rates and the prior.

We perform maximum *a posteriori* inference by maximizing with respect to $q[m]$ (the logarithm of) the conditional probability $P(q[m] | \mathbf{r})$. By Bayes’ rule

$$\log P(q[m] | \mathbf{r}) = \log P(\mathbf{r} | q[m]) + \log P(q[m]) - \log Z \quad (5)$$

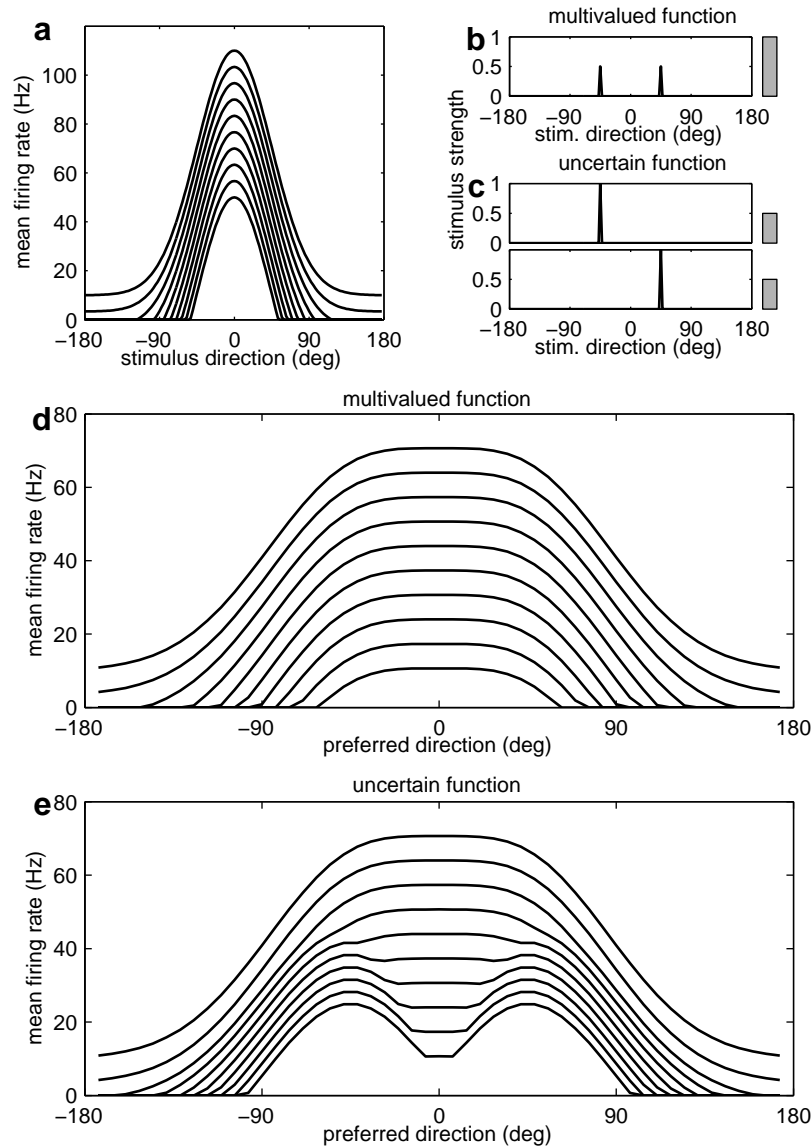


Figure 2: Mean firing rates of cells in a DDPC for multivalued and for uncertain encoded functions. Panel a shows the single-input tuning curves of the 10 neurons in the population with a preferred direction of 0° , but different thresholds. A similar group of neurons is tuned to each of 51 different preferred directions between -180° and 180° . Panels b and c, respectively, show the multivalued and uncertain stimuli to be encoded. The curves on the left in each case represent the different possible functions $m(s)$, while the bars on the right indicate the probabilities of each one. The mean firing rates of the population for the two cases are shown in panels d (multivalued stimulus) and e (uncertain stimulus). Each curve represents the firing rates for the subpopulation with the same threshold.

where the normalizing constant Z does not depend on $q[m]$. Thus, if we held no prior bias with regard to the encoded distribution (*i.e.*, $P(q[m])$ were uniform), this procedure would be a generalization of the maximum-likelihood decoding discussed by Seung and Sompolinsky (1993) for simple population codes. To implement this decoding scheme, we must specify both distributions on the right hand side of equation 5.

The distribution of observed rates $P(\mathbf{r} | q[m])$ (the likelihood) comes from the assumption that activity is measured over a time interval T , during which cell i in the population emits a random number of spikes n_i drawn from a Poisson distribution with mean $\langle \phi_i[m] \rangle_{p[m]} T$. Cells are treated as being statistically independent. The observed rate for the i th cell is thus $r_i = n_i/T$. Given an estimate of the encoded distribution $q[m]$, the term $\log P(\mathbf{r} | q[m])$ can be written in terms of the predicted mean rates $\hat{r}_i = \langle \phi_i[m] \rangle_{q[m]}$ as

$$\log P(\mathbf{r} | q[m]) = \sum_i r_i T \log \hat{r}_i - \hat{r}_i T + \sum_i r_i T \log T - \log(r_i T)!, \quad (6)$$

where we have separated out the terms that do not depend on q .

A convenient choice of prior over the encoded distributions, and one that can be justified by a desire to avoid adding information beyond that encoded, is to make the log prior probability proportional to the entropy of the encoded distribution, *i.e.*,

$$\log P(q[m]) = \alpha H[q] - \log Z_p(\alpha) \quad (7)$$

where $Z_p(\alpha)$ is a normalizing constant. The parameter α sets the importance of the prior relative to the likelihood. Larger values of α lead to decoded distributions with greater uncertainty than would be obtained by maximum-likelihood decoding. For very small α , the prior will have little impact on the likelihood of the decoded value. Instead, it will select between the possible decoded distributions with the same likelihood, choosing the one with the greatest uncertainty.

The encoding rule of equation 4 treats the multiplicity $m(s)$ as a function of a continuous underlying stimulus dimension. To make the implementation computationally feasible in general, we must consider finite (or even better, low) dimensional representations of the functions $m(s)$ and of the distributions over the functions. In this section, we consider discretizations of both of these. This is sensible only in cases where s lives in a small number of dimensions (*e.g.*, one or two). In the next section, we consider mixture models, which are more tightly constrained, but offer the prospect of working in more complicated stimulus spaces.

The first discretization is to approximate the function $m(s)$ by a vector \mathbf{m} , which gives its value at a sufficiently dense grid of points in the stimulus space. This makes the expected values that appear in the predicted rates and in the entropy take the form of integrals over \mathbf{m} . We can then combine the likelihood and prior for the estimated distribution $q(\mathbf{m})$ over the discretized function space (dropping terms that do not depend on q) with a Lagrange multiplier term that enforces the constraint $\int d\mathbf{m} q(\mathbf{m}) = 1$. This leads to a Lagrangian

$$\begin{aligned} \mathcal{L}[q(\mathbf{m})] = \sum_i r_i T \log \int d\mathbf{m} q(\mathbf{m}) \phi_i(\mathbf{m}) - \sum_i \int d\mathbf{m} q(\mathbf{m}) \phi_i(\mathbf{m}) T \\ - \alpha \int d\mathbf{m} q(\mathbf{m}) \log q(\mathbf{m}) - \lambda \left(\int d\mathbf{m} q(\mathbf{m}) - 1 \right). \end{aligned} \quad (8)$$

Stationary points of this functional will correspond to constrained maxima of the posterior distribution of equation 5.

To find those stationary points, we take the variational derivative with respect to $\log q(\mathbf{m})$

$$\frac{\delta \mathcal{L}}{\delta \log q(\mathbf{m})} = q(\mathbf{m}) \left(\sum_i \frac{r_i}{\hat{r}_i} \phi_i(\mathbf{m}) T - \sum_i \phi_i(\mathbf{m}) T - \alpha (1 + \log q(\mathbf{m})) - \lambda \right). \quad (9)$$

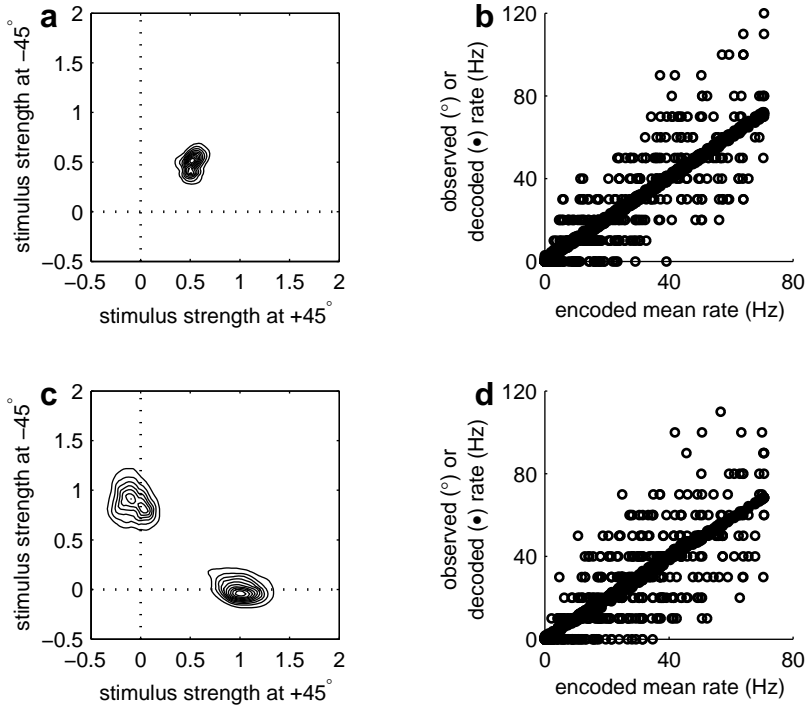


Figure 3: Two-dimensional histogram decoding of multivalued (a–b) and uncertain (c–d) stimuli. The contour plots (a and c) show the decoded estimates in the two-dimensional space of stimulus strength along the $+45^\circ$ and -45° directions. The scatter plots (b and d) show, as a function of the encoded mean rate, the observed rates (open circles – assuming a 100 ms observation time, which leads to the discretization of the observed rates) and the decoded rates (filled circles – see text for explanation).

Setting this equal to 0 and rearranging, we obtain a fixed-point condition that can be used for iterative updates to the estimated distribution:

$$q(\mathbf{m}) = \frac{q(\mathbf{m})}{\sum_i \phi_i(\mathbf{m})} \left(\sum_i \frac{r_i}{\hat{r}_i} \phi_i(\mathbf{m}) - \frac{\alpha}{T} (1 + \log q(\mathbf{m})) - \frac{\lambda}{T} \right). \quad (10)$$

To implement the update, the Lagrange multiplier λ must be chosen so as to result in a properly normalized distribution. This is most easily done by spreading the update over two steps. Given a current estimate $q(\mathbf{m})$, define $\tilde{q}(\mathbf{m}) = q(\mathbf{m}) / \sum_i \phi_i(\mathbf{m})$. First calculate

$$\tilde{q}^{\text{new}}(\mathbf{m}) = \tilde{q}(\mathbf{m}) \left(\sum_i \frac{r_i}{\hat{r}_i} \phi_i(\mathbf{m}) - \frac{\alpha}{T} (1 + \log q(\mathbf{m})) \right), \quad (11)$$

and then normalize

$$q^{\text{new}}(\mathbf{m}) = \tilde{q}^{\text{new}}(\mathbf{m}) + \frac{1 - \int d\mathbf{m} \tilde{q}^{\text{new}}(\mathbf{m})}{\int d\mathbf{m} \tilde{q}(\mathbf{m})} \tilde{q}(\mathbf{m}). \quad (12)$$

In computer implementation, a second discretization is necessary, in which the distribution $q(\mathbf{m})$ is replaced by a discrete approximation on a suitably tiled set of values of \mathbf{m} .

Figure 3 shows the results of decoding noisy versions of the two doubly distributional population codes shown in figure 2, using equations 11 and 12. The mean rates shown in figure 2 were used to generate (Poisson-

distributed) observed counts, assuming a 100 ms observation interval (one might equivalently consider a shorter observation interval and a larger population).

To facilitate computer implementation, the vector m was restricted to two dimensions, representing the two directions of interest in the two example stimulus patterns at $+45^\circ$ and -45° . Thus, the encoded distribution $p(\mathbf{m})$ and the decoded estimate $q(\mathbf{m})$ are both bivariate probability distributions. The decoded estimate $q(\mathbf{m})$ was discretized on a 100×100 grid in the two-dimensional space of stimulus strengths, with values evenly spaced between -0.5 and 2 in each direction. The update equations were iterated 2000 times starting from a uniform initial value of $q(\mathbf{m})$. Figure 3a and c show the final estimates of $q(\mathbf{m})$ for the multivalued and uncertain cases, respectively. Figure 3b and d show the observed rates and the decoded rates (that is, the rates that result if the decoded estimate is re-encoded according to equation 4) for the two populations. Clearly, the decoding procedure is able to identify the correct distribution in both cases, and thus the decoded rates are almost exactly the same as the rates that were originally encoded.

2.3 Decoding to a Mixture

While equations 11 and 12 provide a way to estimate an arbitrary encoded distribution, subject only to the entropic prior, they rapidly become computationally prohibitive as the dimensionality of the underlying distributions increases. Indeed, the size of the discretized histogram parameterization grows exponentially in the dimensionality of the problem. An alternative approach is to assume a more restrictive parameterized form for the estimate of the encoded distribution. Here, we consider the case of a Gaussian mixture in which the component Gaussians have isotropic covariance with a fixed width:

$$q(\mathbf{m}) = \sum_m \pi_m q_m(\mathbf{m}), \quad (13)$$

where

$$q_m(\mathbf{m}) = (2\pi\sigma^2)^{-D/2} e^{-\|\mathbf{m}-\mu_m\|^2/2\sigma^2}. \quad (14)$$

Distributions parameterized in this way lie on a low dimensional manifold within the space of all possible distributions treated above. As is standard with such mixture models, provided the width σ is not allowed to become infinitesimal (and here it will be fixed at a finite value), it is safe to perform maximum likelihood decoding, assuming a flat (and improper) prior over the parameters.

In writing the likelihood for any mixture-based estimate, it is convenient to define the partial estimated rates

$$\hat{r}_{im} = \int d\mathbf{m} q_m(\mathbf{m}) \phi_i(\mathbf{m}) \quad (15)$$

so that the total estimated rate is

$$\hat{r}_i = \sum_m \pi_m \hat{r}_{im}. \quad (16)$$

Then, the likelihood for the mixture model can be written

$$\mathcal{L}(\{\pi_m\}, \{\mu_m\}) = \sum_i r_i T \log \sum_m \pi_m \hat{r}_{im} - \sum_i \sum_m \pi_m \hat{r}_{im} T. \quad (17)$$

It is possible to optimize this likelihood by introducing a Jensen-inequality-derived lower bound, in a similar fashion to the EM algorithm for maximum-likelihood model fitting. However, our experience is that this

scheme converges slowly. An alternative is to optimize the likelihood directly using a conjugate-gradient line search technique. The derivatives with respect to the component means are given by

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}_m} = \sum_i \left(\frac{r_i}{\hat{r}_i} - 1 \right) \pi_m T \frac{\partial \hat{r}_{im}}{\partial \boldsymbol{\mu}_m}. \quad (18)$$

Rather than differentiating with respect to the mixing parameters directly, we reparameterize using a softmax to enforce normalization, $\pi_m = \frac{e^{\alpha_m}}{\sum_{m'} e^{\alpha_{m'}}$, and then differentiate with respect to the new parameters α_m . Since $\frac{\partial \pi_{m'}}{\partial \alpha_m} = -\pi_{m'} \pi_m + \delta_{mm'} \pi_m$, we have

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m} \left(\sum_i r_i T \log \sum_{m'} \pi_{m'} \hat{r}_{im'} - \sum_i \sum_{m'} \pi_{m'} \hat{r}_{im'} T \right) \\ &= \sum_i \sum_{m'} \left(\frac{r_i}{\hat{r}_i} - 1 \right) \hat{r}_{im'} T \frac{\partial \pi_{m'}}{\partial \alpha_m} \\ &= \pi_m \sum_i (r_i - \hat{r}_i) \left(\frac{\hat{r}_{im}}{\hat{r}_i} - 1 \right) T. \end{aligned} \quad (19)$$

Equations 18 and 19 are valid for any mixture approximation to the decoded distribution. To apply them to the mixture of Gaussians in particular, we need an expression for the derivative $\frac{\partial \hat{r}_{im}}{\partial \boldsymbol{\mu}_m}$. To obtain this, we assume a particular form for the encoding function in absence of uncertainty: $\phi_i(\mathbf{m}) = \beta_i [\mathbf{m} \cdot \mathbf{f}_i - T_i]^+$. This corresponds to a threshold non-linearity, which has threshold T_i and a supra-threshold linear response with gain β_i . The vector \mathbf{f}_i is the discretized version of the linear response curve $f_i(s)$. Then, since

$$\hat{r}_{im} = \int d\mathbf{m} \phi_i(\mathbf{m}) (2\pi\sigma^2)^{-D/2} e^{-\|\mathbf{m} - \boldsymbol{\mu}_m\|^2 / 2\sigma^2} \quad (20)$$

we have

$$\frac{\partial \hat{r}_{im}}{\partial \boldsymbol{\mu}_m} = \int d\mathbf{m} \phi_i(\mathbf{m}) (2\pi\sigma^2)^{-D/2} \frac{\mathbf{m} - \boldsymbol{\mu}_m}{\sigma^2} e^{-\|\mathbf{m} - \boldsymbol{\mu}_m\|^2 / 2\sigma^2} \quad (21)$$

$$= \frac{\beta_i}{2} \operatorname{erfc} \left(\frac{T_i - \boldsymbol{\mu}_m \cdot \mathbf{f}_i}{\sqrt{2}\sigma} \right) \mathbf{f}_i. \quad (22)$$

Figure 4 shows the results of decoding the two noisy doubly distributional population codes of figure 2 using the mixture decoding approach. As before, the observed rates used for decoding were Poisson distributed, assuming a 100 ms observation time. The mixture decoding estimates comprised four Gaussian components, each with a fixed isotropic covariance matrix of $0.025I$. The starting point for the decoding had each component with equal weight in the mixture and a random mean. A total of 1000 iterations of conjugate gradient ascent were performed.

Figure 4a and c show the final estimated mixture distributions for the multivalued and uncertain cases respectively. The four curves show the estimated means $\boldsymbol{\mu}_m$ of the four Gaussian component distributions, while the vertical bars to the right indicate the mixing proportions π_m . Figure 4b and d show the observed rates and the decoded rates (that is, the rates that result if the decoded estimate is re-encoded according to equation 4) for the two populations. Clearly, the decoding procedure successfully distinguishes between multiplicity and uncertainty and comes very close to identifying the correct distribution in both cases; the only deviations occurring in the sharing of some stimulus strength between adjacent directions.

2.4 Robustness to noise

The preceding sections have described two algorithms which recover the parameterized distribution most likely to have been encoded given the observed spike-rates for a DDP. The demonstrations of figure 3 and

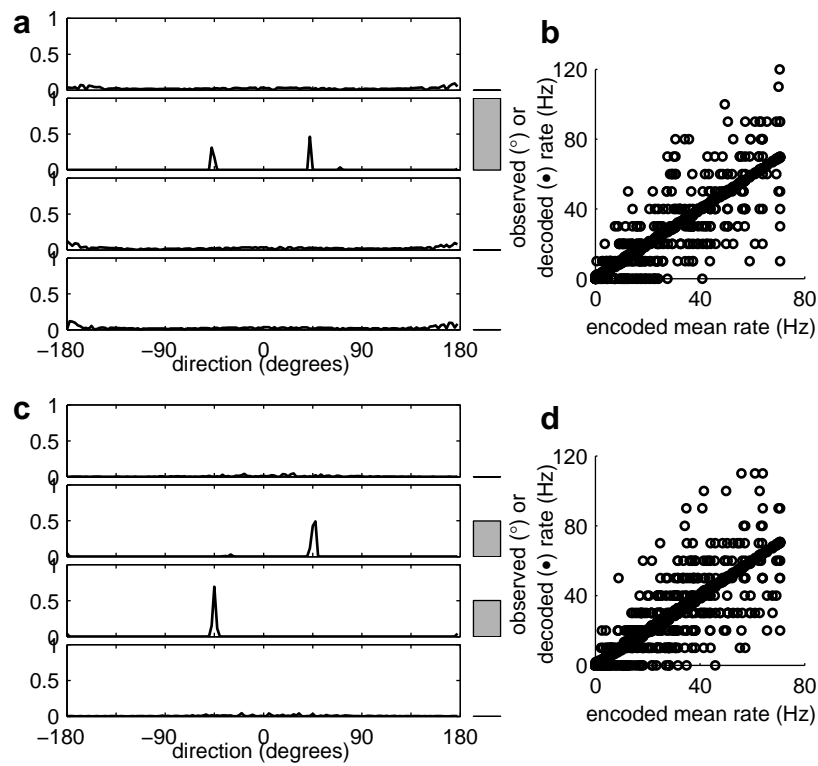


Figure 4: Mixture-model decoding of multivalued (a–b) and uncertain (c–d) stimuli. Panels a and c show the final estimated distributions; the curves on the left show the estimated means μ_m of the four Gaussian component distributions, while the vertical bars to the right indicate the mixing proportions π_m . The scatter plots (b and d) show the observed and decoded rates as a function of the encoded mean rate, with symbol conventions as in figure 3.

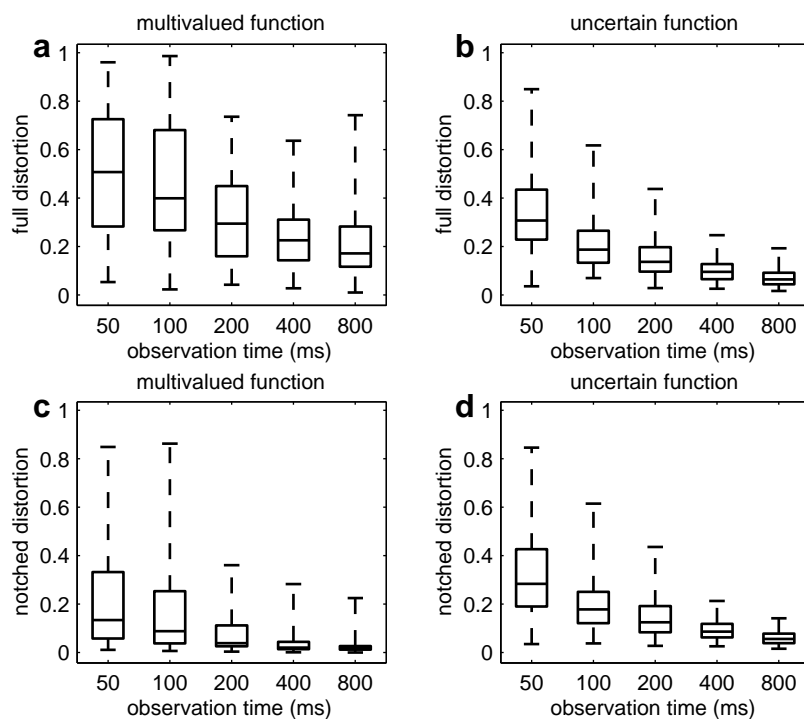


Figure 5: Distributions of observed distortions in the decoding results. Each box-and-whisker plot encapsulates the distribution of distortions obtained when repeatedly decoding noisy DDPC representations of the multivalued (a and c) or uncertain (a and c) functions of figure 2. The horizontal line near the middle of each rectangle indicates the median distortion value; the extent of the rectangle shows the extent of the data that fell within the second and third quartiles of the distribution; while the extent of the whiskers shows the total range of the data. Two different distortion measures were used. Average Euclidean distance between the encoded and decoded function values (“full” distortion) is shown in panels a and b. A “notched” distortion measure that neglected local errors in decoding (see text for details) is shown in c and d.

figure 4 suggest that this maximum *a posteriori* (MAP) decoded distribution can be similar to the encoded distribution, even when the firing rates available to the decoding algorithm have been corrupted by noise. To investigate the coding fidelity of the DDPC more extensively, we carried out a numerical study, the results of which are summarized in figure 5. The complexity of the objects that are encodable using a DDPC – probability distributions over a space of functions – precludes a straightforward analytical treatment of the representational capacity. We therefore approached the question empirically, repeatedly encoding both multivalued and uncertain functions, introducing noise, and then examining the accuracy of the decoded output.

Spike counts were generated repeatedly from Poisson distributions with means given by the DDPC firing rates shown in figure 2 for both the multivalued and the uncertain functions. In all, 500 sets of spike counts were generated for each function, 100 for each of 5 different observation intervals $T \in \{50, 100, 200, 400, 800\}$ ms. Since the distribution of spike counts was Poisson, each doubling of the observation time may alternatively be viewed as a proxy for either a doubling of the population size (replacing each unit by two with the same tuning parameters), or a doubling of the gain of the individual tuning curves. That is, exactly the same effect on fidelity could have been achieved by either of these manipulations.

A MAP estimate of the encoded distribution was recovered from each set of spike counts using the mixture decoding technique of section 2.3. The mixture distributions used in decoding each had four components, with a fixed isotropic covariance matrix of $0.025I$ as in the example of figure 4. For computational efficiency, as well as avoidance of local minima (which are a feature of this particular decoding algorithm, rather than of the DDPC), the means of the Gaussian components were initialized close to the functions that formed the true encoded distributions.

In all cases the MAP estimate correctly identified the number of distinct components that needed appreciable mixing proportion in order to characterize the encoded distribution (that is, 1 for the multivalued distribution and 2 for the uncertain one). The largest mixing proportion assigned in the MAP estimate to any other component was 1.3×10^{-4} . To characterize the accuracy of the decoded means, components with significant mixing proportions in the decoded distribution were matched to the corresponding functions in the encoded distributions, and two distortion measures computed between the component means and the encoded functions. The first (“full”) distortion measure was simply the square-root of the sum of squared differences between the decoded mean and the encoded function (*i.e.*, the Euclidean distance between the vector discretizations of the functions). The overall distortion for the decoded distribution was then the average of the distances obtained for each significant component. The box-and-whisker plots of figure 5a-b summarize the observed distributions of the full distortion in the 100 encodings for each encoded function and observation time. For the uncertain function the median observed distortion drops appreciably at long observation times (or, equivalently, for larger populations). For the multivalued function, however, the decline is less extensive, and the range of observed distortions is not that different to that obtained for much shorter observation times. This was often due to effects such as that seen in figure 4, where some stimulus strength in the MAP estimate is assigned to stimulus directions adjacent to the encoded one. The second (“notched”) distortion measure (similar in spirit to the ϵ -insensitive loss function of Vapnik (1998)), neglected such misassignments, by treating all directions within 7.5° (1 tuning curve center) of the encoded direction as equivalent and measuring only the difference in stimulus strength summed across these directions. The distribution of this “notched” distortion measure is shown in figure 5c-d. Clearly, when local errors are neglected, the distortions are quite small.

3 Discussion

The results of a variety of psychophysical experiments suggest that observers can manipulate both uncertainty and multiplicity in the same perceptual dimensions. However, although abstract perceptual processing algorithms designed to accommodate these results require that uncertainty and multiplicity are represented, our theoretical understanding of information representation in neuronal populations has not previously ac-

commodated this. We have proposed and validated a new scheme, the doubly distributional population code, which is a formal extension of existing methods, and with which both uncertainty and multiplicity can be simultaneously encoded. It might seem hopelessly optimistic to encode anything interesting about an infinite dimensional quantity, namely a probability distribution over functions of a stimulus variable, in the noisy activities of a finite number of neurons. However, we have shown that this can be done appropriately, in some non-trivial cases. We have proven the method using two different decoding algorithms, each with its advantages and disadvantages, but in neither case did we have to use a strong, question-begging prior.

Various points arise about the DDPC, including the provenance of the two-stage scheme for encoding multiplicity and uncertainty, the issue as to how such population codes might support uncertainty-sensitive computations, and possibilities for experimental tests.

3.1 Intuitions underlying the DDPC

The doubly distributional population code suggests a way in which a single population of neurons might distinctly (and, if need be, simultaneously) encode two different types of distributional data: multiplicity and uncertainty. The origin of these two sorts of distribution is quite different, and the different encoding strategies employed by the DDPC reflect this contrast.

Multiplicity is generally associated with the simultaneous activation of multiple sensory neurons (in the simplest cases), or multiple specialized local detector neurons (for example, motion detectors in the early mammalian visual system, or interaural phase-lag detectors in the owl's nucleus laminaris). The outputs of such neurons are pooled (in area MT for motion, or in the inferior colliculus for sound-source location) to form a population representation. For simplicity we and other authors have assumed that these outputs combine roughly linearly, and are then exposed to the spike-rate nonlinearity of cells at the population encoding level. Such a scheme leads immediately to equation 2 (the DPC) and thus also to the DDPC representation of multiplicity.

By contrast, uncertainty is not an inherent property of any particular stimulus, but of the inferences about the world derived by the nervous system on the basis of those stimuli. Sensory experience might be expected to play a significant rôle in shaping neuronal responses to stimuli which are associated with uncertainty; for example, it appears that (generally unambiguous) visually-derived location information is used to guide the auditory localization responses of neurons in the midbrain of the owl (Knudsen and Brainard 1991; Brainard and Knudsen 1993; Hyde and Knudsen 2002). We can gain intuition for the form of the DDPC representation (equation 4) of uncertainty by considering the way in which uncertainty might come to affect population activity through learning. Broadly, a single input comes to be associated with multiple possible outputs, specified by other sources of information, and distributed as $p[m]$. Most reasonable learning schemes tend to average over such distributions, thus leading to equation 4.

More particularly, consider first the case of representation of single-valued quantities, for which the multiplicity functions are just delta functions. In the simplest case, learning is driven by additional information about the state of the world, which we call the teacher signal; this signal may be derived from a different sensory modality, as in the localization system of the owl, or from the same modality at other times. Thus, where the additional information is unambiguous, the teacher signal indicates a single feature value consistent with the sensory input; in response, the population code adjusts its firing rates gradually so that, if the teacher signal remains reliable, it will come to encode that single value without ambiguity. However, in cases where the inference is ambiguous, repeated occurrences of the same input stimulus will be associated with different teacher signals, which will be drawn from the associated uncertainty distribution $p[m]$. As the population learns, then, it modifies its firing rate response to the same input, so as to approach the encoding for first one taught value and then another. Natural learning schemes, including Hebbian learning, make the population settle over time on a response which is the mean of the responses associated with each teacher signal. When the quantity to be encoded is not single-valued, the situation is similar. In this case the representations for both multiplicity and uncertainty must be learned. However, now, each teacher signal must reflect an

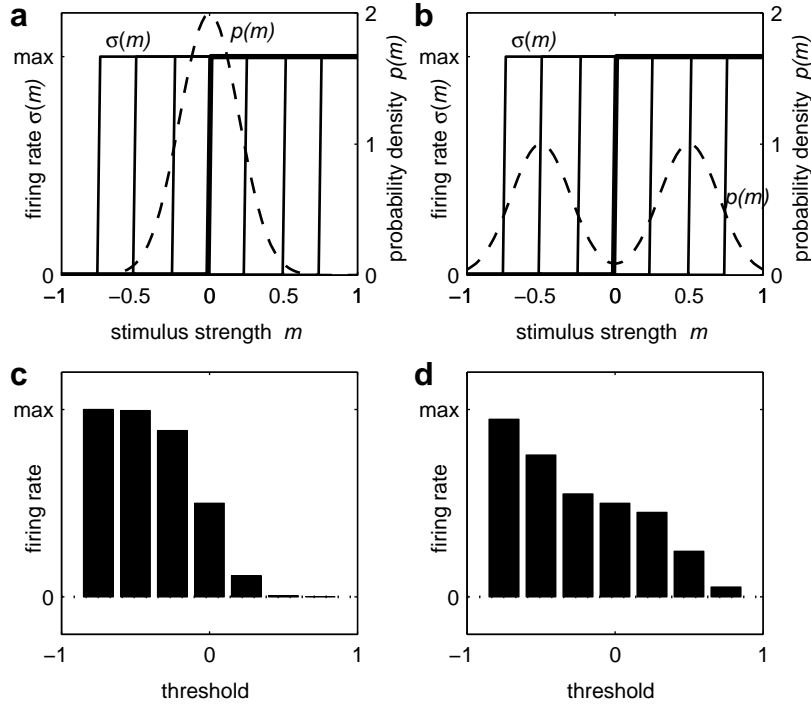


Figure 6: Encoding of a one-dimensional uncertainty distribution. Panels **a–b** show two different probability distributions $p(m)$ (dashed lines), along with a family of 7 step-function non-linearities $\sigma(m)$ (solid lines; a single step function is shown in heavier ink for clarity). Panels **c–d** show the mean firing rates under the DDPC for a group of neurons with the different non-linearities of **a–b**, as a function of their thresholds. The rates in **c** correspond to the unimodal distribution in **a**, while those in **d** reflect the bimodal distribution of **b**.

instance of the entire multi-valued inference associated with the stimulus (for example, the visually-derived location of all of the sound sources). The result will be the same as before: the firing rates learned by the system are the means of the firing rates required to represent each example multiplicity function. This leads exactly to the DDPC equation 4.

Such observations provide an intuitive justification of the form of the DDPC, but do not explain its success in encoding the two types of distributional data. The representation of multiplicity is, of course, identical to that used in the DPC, and so has been treated previously. Briefly, and in the language of the discretized code, each linear encoding function defines a basis vector in the space of multiplicity vectors \mathbf{m} ; the neuronal firing rates then report non-linear transforms of the dot-products between the encoded multiplicity vector and this generally over-complete basis. More interesting is the novel representation of uncertainty proposed here. Some insight into this representation can be gained by considering the encoding of a probability distribution along one of these basis vectors. The stimulus strength parallel to this basis element is a scalar quantity which we denote by m ; we then seek to encode a one-dimensional probability distribution $p(m)$ (a marginal of the full distribution $p(\mathbf{m})$). For simplicity, consider a group of neurons with identical linear encoding functions, all aligned with the chosen basis vector, and Heaviside function non-linearities with different thresholds (that is, the firing rate of the i th neuron for an unambiguous input is zero when the input magnitude falls below some threshold, T_i , and equal to a fixed non-zero level, r_i^{\max} , otherwise). If we set all scale factors (such as the norm of the linear encoding functions) to unity, then the firing rates under the uncertainty distribution $p(m)$ are

$$r_i(p(m)) = \int dm p(m) \sigma_i(m) = r_i^{\max} \int_{T_i}^{\infty} dm p(m), \quad (23)$$

and so, taken together, they give a discretized version of the complementary cumulative distribution of $p(m)$. This situation is depicted in figure 6 for two different uncertainty distributions. Clearly, the encoding cells distinguish between the two cases in their firing rates. The situation is more complex for different non-linearities and for non-parallel linear encoding functions, but this basic intuition still underpins the successful representation of uncertainty.

3.2 Computations with uncertainty

Our explicit demonstration of decoding proves that DDPCs can represent uncertainty and multiplicity. However, the brain seems to use multiple stages of population codes, and so the key biological computation is not likely to be the sort of decoding that we performed, but rather population-coded computations. One paradigm case of this is integrating multiple sources of information. In our explanatory examples such as the aperture problem, we have mostly considered the importance of representing and manipulating uncertainty correctly in order to integrate information from within sensory submodalities such as visual motion. However, the class of neural computations for which there is the most evidence for the use of uncertainties is cross-modal cue integration (*e.g.*, Ernst and Banks 2002).

Deneve *et al.* (2001) have described how an appropriately designed attractor network could carry out such integration in a maximum *a posteriori* manner when single-valued unimodal inferences are represented by simple population codes whose individual uncertainties are represented by the magnitudes of their initial activities, via an assumption about Poisson encoding noise. Deneve *et al.* show that their network settles to an attractor state corresponding to an integrated inference that is close to the statistically optimal value. This impressive result shows a neurally plausible scheme for *computing* with correctly represented uncertainties. However, their coding choices impose strict limitations on the sorts of distributions that can be integrated efficiently. In particular, the attractor structure of their network permits only a single fixed-width hill of activity in the population, and thus disallows multiplicity in the result of the integration. Also, there are problems with the magnitude code for uncertainty, which have previously been investigated by Zemel *et al.* (1998).

An obvious direction for future work is to extend Deneve *et al.*'s attractor-based integration scheme to the more complete representational framework described here. Crudely, instead of having attractors that represent a single value of an underlying multidimensional stimulus \mathbf{s} , the choice of which is determined by uncertainty-sensitive integration across the multiple dimensions, one might consider attractors that represent single functions of the unidimensional components of \mathbf{s} , namely $m(s)$, whose collective selection is determined in the same way. Key to Deneve *et al.*'s scheme are joint basis functions in the multiple dimensions of \mathbf{s} ; our new scheme would require joint basis functions in the multiplicity functions themselves. Done naively, this would lead to an inordinate explosion of the number of basis functions; it would be essential to consider ways of preventing this, for instance, by using hierarchies of basis functions.

3.3 Experimental Issues

Our proposal of the DDPC has largely been driven by the mismatch between the information demands of perceptual processing algorithms implied by the results of behavioural experiments, and the current theoretical view of information representation in neuronal populations. Many experiments suggest that observers manipulate both uncertainty and multiplicity about the same quantities, and the DDPC is the first neural coding scheme to be proposed for which both these entities can be encoded simultaneously. The specific predictions made by this proposal must now be examined in physiological experiments. We consider here the types of experimental test that might be appropriate.

Experimental tests of all neural coding hypotheses are complicated by the fact that the question of *how* something is encoded in neuronal activities is inextricable from the question of *what* is encoded. Codes such

as the DDPC, which describe a representation for inferential uncertainty, are particularly sensitive to subtle misidentifications of the encoded quantity. For example, neurons sensitive to visual motion are found in many areas of the neocortex, such as V1, MT and MST in primates. It is possible that cells in all of these areas encode essentially the same features of visual motion, albeit at different spatial scales. Alternatively, it might be that V1 cells encode local motion energy, while MT (or possibly a portion of MST) encodes the motion of objects in the visual scene inferred from these local motion energy signals (see, for example, Pack and Born 2001). Uncertainty in these two different motion-related quantities would arise in very different ways; local motion energy estimates are affected only by receptor and neural transmission noise, while inferences about object motion also reflect uncertainty due to the ill-posed nature of the inference, as illustrated by the aperture and correspondence problems. Thus, depending on which of these aspects of visual motion is represented, radically different degrees and forms of uncertainty would be expected.

A prerequisite for testing the DDPC predictions for (say) visual object motion encoding must therefore be to identify the stage in the visual system at which the results of inference about object motion are represented. One standard way to do this is to use stimuli which induce similar object motion percepts but are physically distinct. Similar responses to the stimuli would indicate that object motion, rather than other physical attributes of the stimuli, is being encoded. Then the encoding of inferential uncertainty could be probed with stimuli that invoked similar distributions of uncertainty by different mechanisms, for example, exploiting the correspondence problem in one case and the aperture problem in the other. If the firing rates of a population of motion-sensitive cells were found to be similar in these two cases, and different from those invoked by a third stimulus which invoked a percept of multiplicity with a pattern similar to the uncertainty, we would have a strong indication that the appropriate encoding layer had been found.

While seemingly straightforward, the search for a representation of uncertainty in the sustained spiking of a population of neurons might be complicated somewhat by the interplay of representation and computation within the nervous system. For example, it is possible that a population code for perceptual uncertainty might be only transiently visible, subsequently being overwritten through feedback connections. This happens, for instance, in the model described by Deneve *et al.* (and discussed above), where populations that represent uncertain unimodal information are embedded within a multimodal attractor network. As the network activity evolves, the uncertain single-modality-driven activities in these populations are replaced with the cross-modal integrated value, which may contain little or no uncertainty. Even in this case however, the initial activities must represent the uncertainties correctly, as must the feedforward input to the recurrently coupled attractor network.

Once a population encoding uncertainty, either transiently or persistently, has been located, testing the particular predictions of the DDPC becomes straightforward. The encoding equation (4) describes a particular relationship between the response to single-valued, certain motion (given by $\sigma(f_i(s))$) and the activities induced by patterns of multiplicity or uncertainty. Comparison of this predicted relationship to the observed activities in the various cases would reveal whether the DDPC does indeed provide an accurate account of population coding of inferred quantities.

References

- Ashby, F. G., Prinzmetal, W., Ivry, R., and Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychol. Rev.* 103(1), 165–192.
- Atkins, J. E., Fiser, J., and Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Res.* 41(4), 449–461.
- Baldi, P. and Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biol. Cybern.* 59(4–5), 313–318.
- Brainard, M. S. and Knudsen, E. I. (1993). Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl. *J. Neurosci.* 13(11), 4589–4608.

- Dayan, P. and Abbott, L. (2001). *Theoretical Neuroscience*. Cambridge, MA: MIT Press.
- Deneve, S., Latham, P. E., and Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* 4(8), 826–831.
- Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870), 429–433.
- Eurich, C. W. and Wilke, S. D. (2000). Multidimensional encoding strategy of spiking neurons. *Neural Comput.* 12(7), 1519–1529.
- Feng, A. S. and Ratnam, R. (2000). Neural basis of hearing in real-world situations. *Annu. Rev. Psychol.* 51, 699–725.
- Georgopoulos, A. P., Schwartz, A. B., and Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science* 233(4771), 1416–1419.
- Hinton, G. E. (1984). Distributed representations. Technical Report CMU-CS-84-157, Computer Science Department, Carnegie-Mellon University.
- Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence* 17(1–3), 185–203.
- Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195(1), 215–243.
- Hyde, P. S. and Knudsen, E. I. (2002). The optic tectum controls visually guided adaptive plasticity in the owl’s auditory space map. *Nature* 415(6867), 73–76.
- Knudsen, E. I. and Brainard, M. S. (1991). Visual instruction of the neural map of auditory space in the developing optic tectum. *Science* 253(5015), 85–87.
- Pack, C. C. and Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area mt of macaque brain. *Nature* 409(6823), 1040–2.
- Paradiso, M. A. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybern.* 58(1), 35–49.
- Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. *Nat. Rev. Neurosci.* 1(2), 125–132.
- Pouget, A., Deneve, S., Ducom, J. C., and Latham, P. E. (1999). Narrow versus wide tuning curves: What’s best for a population code? *Neural Comput.* 11(1), 85–90.
- Ramachandran, V. S. and Anstis, S. M. (1985). Perceptual organization in multistable apparent motion. *Perception* 14(2), 135–143.
- Saberi, K., Takahashi, Y., Farahbod, H., and Konishi, M. (1999). Neural bases of an auditory illusion and its elimination in owls. *Nat. Neurosci.* 2(7), 656–659.
- Saunders, J. A. and Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Res.* 41(24), 3163–3183.
- Seung, H. S. and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proc. Natl. Acad. Sci. USA* 90, 10749–10753.
- Snippe, H. P. and Koenderink, J. J. (1992). Information in channel-coded systems: correlated receivers. *Biol. Cybern.* 67(2), 183–190.
- Takahashi, T. T. and Keller, C. H. (1994). Representation of multiple sound sources in the owl’s auditory space map. *J. Neurosci.* 14(8), 4780–4793.
- Treue, S., Hol, K., and Rauber, H. J. (2000). Seeing multiple directions of motion — physiology and psychophysics. *Nat. Neurosci.* 3(3), 270–276.
- Vapnik, V. (1998). *Statistical Learning Theory*. New York: Wiley.
- Weiss, Y. and Adelson, E. H. (1998). Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision. AI Memo 1624, MIT.

- Wilke, S. D. and Eurich, C. W. (2002). Representational accuracy of stochastic neural populations. *Neural Comput.* *14*(1), 155–89.
- Zemel, R. S. and Dayan, P. (2000). Distributional population codes and multiple motion models. In *Advances in Neural Information Processing Systems*, Volume 11. MIT Press.
- Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Comput.* *10*(2), 403–430.
- Zhang, K. and Sejnowski, T. J. (1999). Neuronal tuning: To sharpen or broaden? *Neural Comput.* *11*(1), 75–84.