RESEARCH ARTICLE

Efficient sensory coding of multidimensional stimuli

Thomas E. Yerxa [1¤*], **Eric Kee** [2], **Michael R. DeWeese** [1,3,4], **Emily A. Cooper** [3,5]

1 Department of Physics, University of California, Berkeley, Berkeley, California, United States of America, **2** Uber Advanced Technologies Group, Pittsburgh, Pennsylvania, United States of America, **3** Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, California, United States of America, **4** Redwood Center for Theoretical Neuroscience, University of California, Berkeley, Berkeley, California, United States of America, **5** School of Optometry & Vision Science Program, University of California, Berkeley, Berkeley, California, United States of America

¤ Current address: Center for Neural Science, New York University, New York, New York, United States of America
* tey214@nyu.edu

Abstract

According to the efficient coding hypothesis, sensory systems are adapted to maximize their ability to encode information about the environment. Sensory neurons play a key role in encoding by selectively modulating their firing rate for a subset of all possible stimuli. This pattern of modulation is often summarized via a tuning curve. The optimally efficient distribution of tuning curves has been calculated in variety of ways for one-dimensional (1-D) stimuli. However, many sensory neurons encode multiple stimulus dimensions simultaneously. It remains unclear how applicable existing models of 1-D tuning curves are for neurons tuned across multiple dimensions. We describe a mathematical generalization that builds on prior work in 1-D to predict optimally efficient multidimensional tuning curves. Our results have implications for interpreting observed properties of neuronal populations. For example, our results suggest that not all tuning curve attributes (such as gain and bandwidth) are equally useful for evaluating the encoding efficiency of a population.

Author summary

Our brains are tasked with processing a wide range of sensory inputs from the world around us. Natural sensory inputs are often complex and composed of multiple distinctive features (for example, an object may be characterized by its size, shape, color, and weight). Many neurons in the brain play a role in encoding multiple features, or dimensions, of sensory stimuli. Here, we employ the computational technique of population modeling to examine how groups of neurons in the brain can optimally encode multiple dimensions of sensory stimuli. This work provides predictions for theory-driven experiments that can leverage emerging high-throughput neural recording tools to characterize the properties of neuronal populations in response to complex natural stimuli.

Introduction

A functioning nervous system is metabolically expensive, but offers a significant advantage in evolutionary fitness. This observation motivates the idea that natural selection has acted to optimize the nervous system subject to biological constraints. Indeed, optimality has proven to be a powerful concept for explaining the function of neural systems, circuits, and individual cells. But optimality arguments come with their own set of auxiliary questions. Optimal in what sense? Subject to what specific constraints?

A common framework for choosing these criteria is the *efficient coding hypothesis*. This hypothesis posits that sensory neurons maximize the information that they encode about the world while using as few resources (e.g., action potentials) as possible [1, 2]. For example, by defining a set of neurons as optimal if they may be used in linear combination to accurately represent an arbitrary image, and imposing the constraint that for any given input as little neural activity should be used as possible, we arrive at a sparse coding model [3]. Considering instead the constraint that each neuron should encode as little redundant information as possible gives rise to independent components analysis [4, 5].

To understand how neuronal populations can operate efficiently during natural behavior, it is often advantageous to consider that natural sensory signals have robust statistical regularities (See [6] and [7] for review). Thus, neurons are generally tasked with encoding a small subset of all possible stimuli. Said another way, the probability distribution of stimuli experienced in the natural environment is highly non-uniform. Given a particular non-uniform probability distribution of a 1-D stimulus, a body of recent work considers how to optimize the efficiency of neuronal populations in which each neuron is characterized by a tuning curve that defines its mean response as a function of the stimulus value [8–14]. This simple population model yields several appealing solutions for optimal encoding populations with testable predictions. For example, a recent formulation predicts that neurons tuned to more frequently occurring stimulus values should be more densely packed and tightly tuned, but should not necessarily have higher or lower response gain (Fig 1, upper row) [8, 9]. These predictions with respect to density and tuning shape can successfully explain empirical measurements in a variety of sensory neuron populations [15–17].

While 1-D population models have yielded a range of insights and predictions for neural efficiency, it is important to note that sensory neurons throughout the brain are selective to more than one stimulus dimension. Canonical examples include neurons in mammalian primary visual and auditory cortices that can jointly encode spatial frequency/orientation and pitch/timbre, respectively [18, 19], as well as medial temporal neurons that encode multiple variables relevant to 3-D spatial position or pose [20–23]. In higher-level sensory cortices, larger dimensionality in the encoding regimes may be common (e.g., [24, 25]). To date, the pay off of the efficient coding framework for understanding high-dimensional stimulus encoding has been more modest because fewer predictions for efficient multidimensional tuning curves are formalized [13, 26]. As a consequence, it is more challenging to predict and evaluate efficiency in higher dimensions. With high-throughput experimental paradigms and large data sets becoming more common in neuroscience, the ability to model optimal encoding of higher dimensional stimuli is increasingly relevant.

In the current work, we make advances towards filling this gap by extending existing frameworks for modeling efficient 1-D population codes to higher dimensions (summarized in Fig 1, bottom row). We first review the formulation of an encoding model that employs a constrained heterogeneous population of neurons to optimally (in the information preserving sense) encode a 1-D stimulus using a finite and fixed number of spikes [8, 9]. As our first contribution, we generalize this formulation by re-parameterizing the heterogeneous population
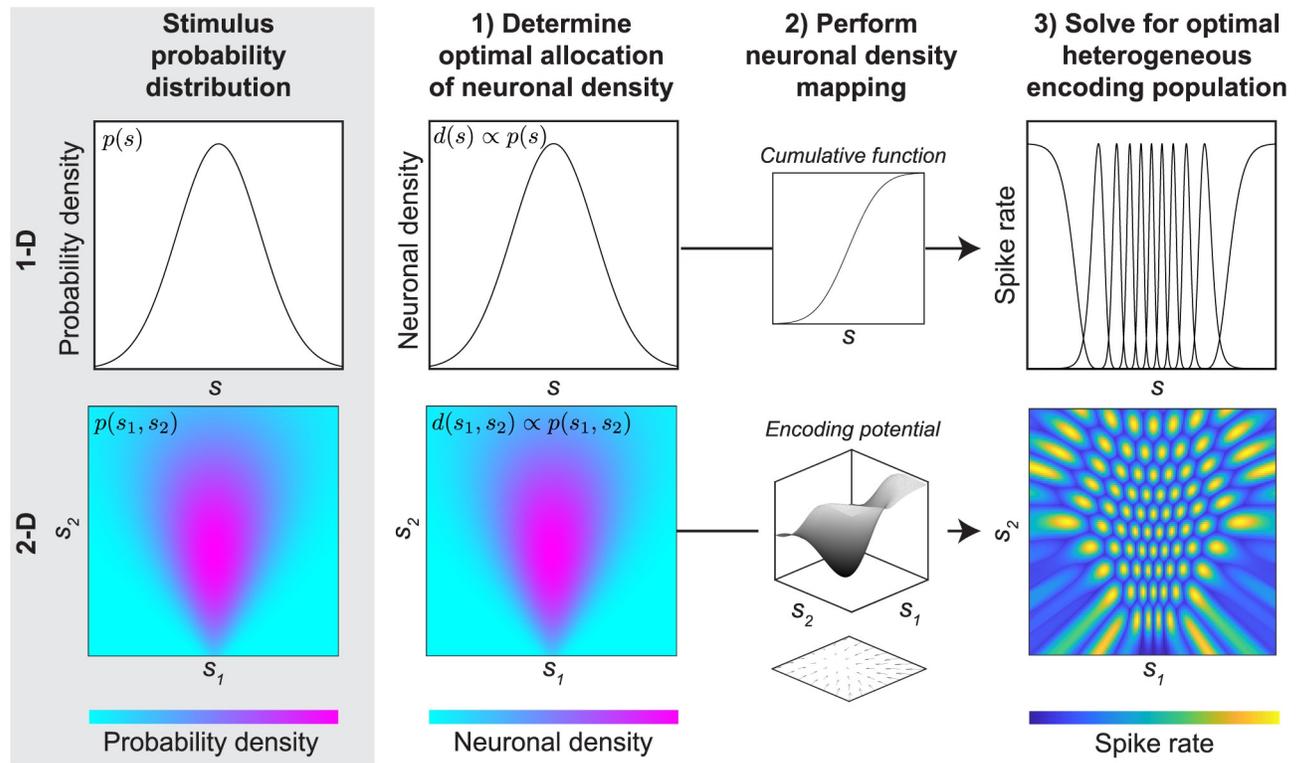
Fig 1. Overview of optimally efficient heterogeneous neuronal encoding populations in 1-D and 2-D. Top row: Based on prior work [8, 9], we derive a closed-form solution for an optimal heterogeneous neuronal population encoding a non-uniform 1-D stimulus probability distribution. This population has a neuronal density that is proportional to the probability of the stimulus. The neuronal density can be mapped to a specific population (here, represented by a set of Gaussian tuning curves) by using the cumulative of the probability distribution $p(s)$ to warp the stimulus coordinates (s) over which each neuron's tuning curve is defined. The resulting heterogeneous 1-D neuronal population is compressed in regions of higher probability and expanded in regions of lower probability. Bottom row: Neurons throughout the nervous system encode more than one stimulus dimension. Here, we extend the previous 1-D framework to examine optimally efficient neuronal populations in arbitrary dimensions. In closed form, we show that given the same assumptions used for 1-D populations, higher dimensional neuronal populations should also have density proportional to probability. However, in higher dimensions with statistical dependencies, density remapping cannot be achieved via the cumulative. Instead, we show that in both 1-D and higher dimensions, the optimal mapping is embodied by the gradient of a scalar potential function that reflects the displacement necessary to uniformly distribute neuronal density as a function of probability. This encoding potential is illustrated here as a surface, with its inverse gradient illustrated as a vector field below. The vector field is inverted to provide a more intuitive visual of how neuronal density gets condensed to areas of high probability. This gradient can be numerically optimized in 2-D for a given density function, which we exploit to derive optimal 2-D neuronal populations.

https://doi.org/10.1371/journal.pcbi.1008146.g001

in terms of a more generic transformation, based on recent work describing transformations between probability distributions [27]. This generalization replicates prior work in 1-D and enables a natural extension to solve for the parameters of optimal encoding populations in higher dimensions. We show that this extension produces similarly elegant predictions for the relationship between stimulus probability and neuronal density in populations encoding an arbitrary number of stimulus dimensions. However, predicting the actual shapes of multidimensional tuning curves when there are statistical dependencies between stimulus dimensions remains more complicated. Focusing on 2-D stimuli, we show that it is nonetheless possible to utilize a numerical method (developed in [27]) to predict tuning curves of efficient encoding populations across the stimulus space. Using this approach, we examine the predicted tuning properties for neuronal encoding populations optimized for 2-D stimuli. These results highlight both the utility and limitations of simple encoding models for evaluating the efficiency of neuronal populations that encode multiple stimulus dimensions.

Methods

Encoding objective

One natural question in the context of neural coding is: how much information about a stimulus is contained in the responses of a population of sensory neurons? The average amount of information learned about some random variable (say, the value of a stimulus) by observing another random variable (say, the firing rate of a sensory neuron) is given by the mutual information ($\mathcal{M}[x, y]$) [28]:

$$\mathcal{M}[x, y] = \int \int p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) dx dy. \tag{1}$$

Here, x and y denote any two random variables, $p(\cdot)$ denotes the probability distribution of a variable, and $p(x, y)$ denotes their joint probability. Though mutual information appears to be a strong candidate for an objective function to maximize for efficiency, it is difficult to optimize because doing so requires integration over all variables. To avoid this issue, researchers often instead optimize approximations of the mutual information that make use of a quantity called the Fisher information ($\mathcal{F}$) [9, 26, 28–30]. The Fisher information provides another way to quantify the amount of information an observable random variable (x) carries about a parameter (y). It can be computed from the conditional probability distribution of x given y as

$$\mathcal{F}(y) = E_y\left[-\frac{\partial^2 \log(p(x|y))}{\partial y^2}\right], \tag{2}$$

where $E_y[\cdot]$ denotes taking the expectation of the bracketed property over the parameter y. When the curvature of the conditional distribution, $p(x|y)$, is high, small changes in the parameter cause large changes in the probability. Thus, observations in this region are more informative.

The Fisher information limits the variance with which a parameter may be estimated (through a relationship known as the Cramer-Rao bound). This fact has been used to derive an approximation of mutual information [11, 12, 31]:

$$\mathcal{M}[x, y] \geq \mathcal{H}(y) + \int p(y) \frac{1}{2} \log\left(\frac{\mathcal{F}(y)}{2\pi e}\right) dy. \tag{3}$$

Here $\mathcal{H}(y)$ is the entropy of the probability distribution for the parameter y, and is a constant. For an unbiased estimator, this approximation approaches equality in the limit of low noise, which occurs in the context of neural coding when there are either a large number of neurons or a large number of action potentials fired. A recent and detailed exploration of Fisher information and mutual information can be found in [12]. Following previous work, we will thus use the Fisher information in the neuronal population response to the sensory stimulus to determine the quantity to be maximized in our neuronal population model [8–11, 13, 14, 31–33].

Neuronal population model

The encoding characteristics of sensory neurons are often described via a tuning curve, which is a function that defines how an idealized neuron responds to different values of a particular stimulus. In this work, we will consider neuronal responses in the form of rate codes, where the tuning curve of a neuron defines the mean firing rate as a function of the presented stimulus value. There is an inherent trade off between the range of stimulus values that an individual neuron can encode and how much information it can convey. A very broad tuning curve would be able to respond to a large number of stimulus values, but because the magnitude of

the response changes slowly it will be more difficult to discriminate between different encoded values. Conversely a very narrow tuning curve might be able to encode a single value with very high certainty, but would leave a large set of values unaccounted for. In practice, sensory systems often use a large population of neurons (each with their own, different tuning curve) to encode the same sensory variable. This strategy is known as population coding.

We will consider a neuronal population in which each neuron's firing rate is determined by an independent Poisson process. In describing this population, we adopt notation from [9], as our work builds on theirs. A (homogeneous) Poisson process is a model for a series of independent events (in this case, action potentials) in which events are equally likely to occur at any time independent of the number or timing of previous events. The distribution that describes the probability of observing r events within a given time interval for a Poisson process is defined by a single parameter, λ, which represents the expected (mean) number of events in the interval. The probability distribution for observing r events for a Poisson process with mean λ is:

$$p(r) = \frac{\lambda^r e^{-\lambda}}{r!}. \tag{4}$$

In the context of neural coding, the parameter λ that determines the mean firing rate r of a neuron in response to a stimulus is defined by that neuron's tuning curve $h(\cdot)$. Since neurons can be tuned over multiple stimulus dimensions, we can denote the stimulus attributes that modulate the firing rate with the vector $\boldsymbol{s}$ of any dimensionality. However, we will initially be focusing on the case of 1-D stimuli, so we will denote the 1-D stimulus as s and the tuning curve as $h(s)$. Thus, the probability of a sensory neuron firing with rate r conditioned on the presentation of the stimulus s is:

$$p(r|s) = \frac{h(s)^r e^{-h(s)}}{r!}. \tag{5}$$

We will consider a population of N neurons, each with its own tuning curve, which respond to the presented stimulus (conditionally) independently of all of the other neurons in the population. Thus, the probability of observing a population response $\boldsymbol{r}$ upon the presentation of stimulus s is:

$$p(\boldsymbol{r}|s) = \prod_n^N \frac{h_n(s)^{r_n} e^{-h_n(s)}}{r_n!}. \tag{6}$$

Here n indexes the population of neurons, h_n is the tuning curve of the nth neuron, and $\boldsymbol{r}$ is an N-dimensional continuous vector whose elements (r_n) denote the firing rate of the nth neuron in response to some stimulus. Because each neuron is assumed to act conditionally independently of any other neuron in the population, the joint probability of the responses is modeled as the product of the individual responses.

The primary goal of this work will be to find how a population of sensory neurons should respond to the presentation of a stimulus in order to maximize the mutual information between the stimulus and response. To do so we will optimize an approximation of $\mathcal{M}[\boldsymbol{r}, s]$. As such, we will use the Fisher information for the independent Poisson population encoding model. As described in [33], this quantity is given by:

$$\mathcal{F}(s) = \sum_n^N \frac{h_n'^2(s)}{h_n(s)}, \tag{7}$$

where $h_n'(s)$ denotes the first derivative of the tuning curve.

Parameterization of 1-D heterogeneous neuronal populations

Our parameterization of hetereogeneous populations follows on the formulation introduced in [8, 9]. As a starting point (Fig 2), we will define a population of 1-D tuning curves, where each curve has an identical shape determined by the function $h(s)$, and individual tuning curves ($h_n(s)$) fall at evenly spaced intervals so that $h_n(s) = h(s - s_n)$. Here, s_n denotes evenly spaced values that determine the preferred stimulus for each neuron (Fig 2A). Additionally, we assume that the population of neurons approximately tiles the stimulus space (tuning curves are thus not separated by large gaps, nor do they take the form of monotonically increasing functions) [8, 9, 34, 35]. Specifically,

$$\sum_n^N h_n(s) \approx k. \tag{8}$$

This means that the number of spikes emitted by the population is an approximately constant (k) function of the presented stimulus. The constant spike rate means that the Fisher information for this uniform population (which we denote $\mathcal{F}_u(s)$) is also approximately constant (Eq (7), Fig 2B). Therefore, this population is non-optimal for encoding stimuli with non-uniform probability distributions.
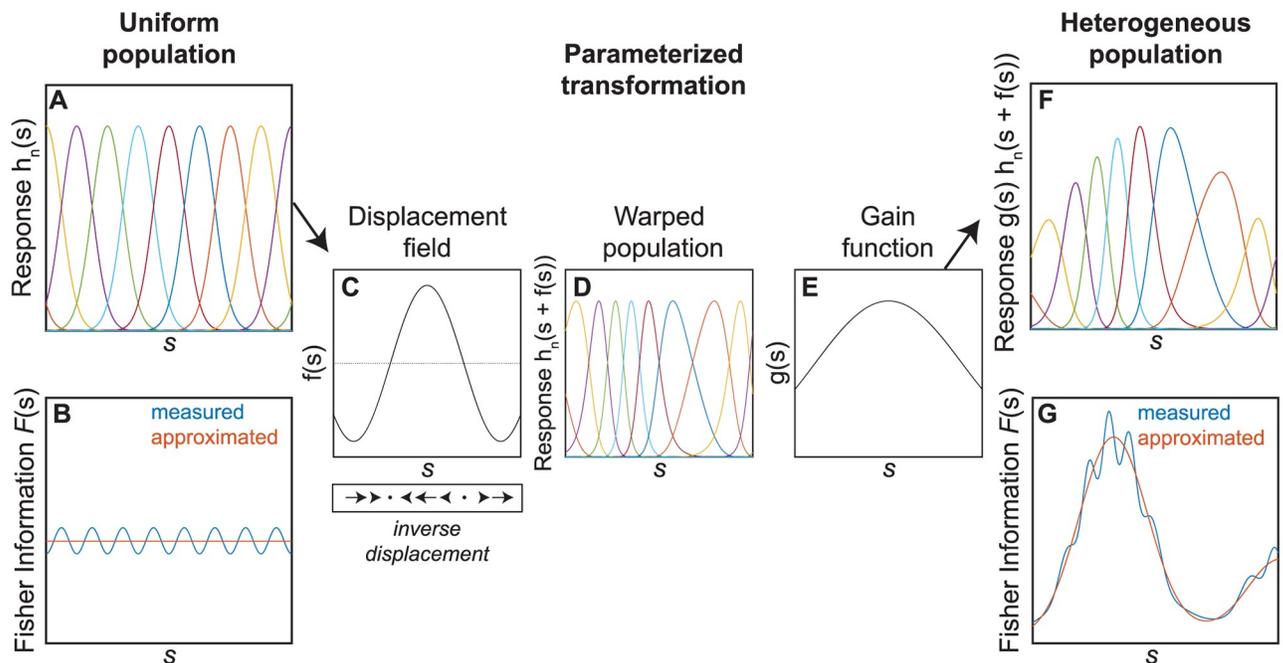

Fig 2. Parameterizing hetereogeneous neuronal populations. A) A uniform population of neurons approximately tiles the stimulus space (s) with identical, equally spaced Gaussian tuning curves. B) The Fisher information of this population is roughly uniform (blue line), matching the approximation in Eq (16) (red line). C) A displacement field that is smooth and slowly varying relative to the tuning curves. These displacement values apply to the stimulus space, arrows below illustrate the direction and magnitude of shifts in the resulting tuning curves defined over s (which corresponds to the inverse of the displacement field). D) After the displacement field is applied, the neuronal population now has heterogeneous tuning curves. Displacements that stretch the stimulus space result in denser, narrower tunings. Displacements that compress the stimulus space result in sparser, wider tunings. E) A gain function that is smooth relative to the tuning curves can also allow neurons to have different response magnitudes. F) Following the application of both the displacement field and the gain function, we have a transformed heterogeneous population with variable tuning curves. G) The Fisher information in the hetereogenous population is no longer uniform, as illustrated by the measured (blue) and approximated (red) lines. S1 Fig illustrates the consequences when the displacement field and gain function are not smooth and slowly varying with respect to the tuning curves.

To determine an optimal encoding population for a non-uniformly distributed stimulus, we will define two functions that warp and scale this uniform population, and thus reallocate Fisher information to different regions of the stimulus space. The choice of parameters and the constraints on their smoothness build on the formulation introduced in [8, 9]. However, it is important to note that alternative parameterizations can produce quite different predictions [12, 30], which we will consider in the Discussion.

Specifically, we will parameterize the population with two smooth and continuous functions. The first is a displacement field, $f(s)$, which will warp the tuning curves to change their local density (Fig 2C and 2D). The second is a gain function, $g(s)$, which will control the maximum response of neurons to a given stimulus (by multiplicatively scaling the size of the response) (Fig 2E and 2F). Each of these functions will also be constrained to be slowly changing relative to the individual tuning curves. This property of each parameter will force the warped and scaled tuning curves to (approximately) retain their functional form and tile the space (S1 Fig illustrates the consequences of violating this constraint).

The notable difference between this formulation and that of previous work [8, 9] is that we use a displacement field to change neuronal density. The displacement field can be thought of as transforming the original stimulus space (which is simply the value of some stimulus variable s) to a new domain as follows:

$$\hat{s} = s + f(s). \tag{9}$$

Following the nomenclature suggested by [30], we will refer to the original space as the stimulus space (which is the space defined by the stimulus variable s), and the transformed space as the sensory space ($\hat{s}$, which is the space defined by some transformation as denoted in Eq (9)). In contrast to this approach, previous work changed neuronal density by applying the cumulative of a function that directly described the desired density of tuning curves. Shifting the population based on the cumulative of the desired density function has the elegant property of producing the desired density directly [8, 9, 30, 36]. However, the reliance on a cumulative function is not appropriate when considering higher dimensions with statistical dependencies, because the multivariate cumulative treats each dimension separately. The more general displacement field is still related to the density function, and in 1-D there is still a one-to-one relationship between displacement and density.

What is the relationship between the transformation in Eq (9) and density in 1-D? The first derivative of a 1-D transformation directly embodies the change in density associated with that transformation (indeed, this is why the cumulative of the desired density function was used in prior work). Thus, if the density with which the tuning curves tile the space is defined by $d(s)$, then in 1-D we simply take the first derivative of Eq (9) to obtain the density after transformation:

$$d(s) = 1 + f'(s). \tag{10}$$

Intuitively, this relationship arises because $f(s)$ defines how much each point on each tuning curve is shifted in a given direction, and successive points being shifted in the same direction by an increasing amount will become less densely packed. Similarly, successive points being shifted in the same direction by a decreasing amount will become more densely packed.

Fisher information in the 1-D heterogeneous neuronal populations

The displacement field and gain function each have a distinct effect on the Fisher information of the population. The effect of the displacement field on the Fisher information is:

$$\mathcal{F}_f(s) = \mathcal{F}_u(\hat{s})\left(\frac{d\hat{s}}{ds}\right)^2. \tag{11}$$

Here, $\mathcal{F}_f(s)$ denotes the Fisher information after the stimulus space has been 'warped' into the sensory space. Because $\mathcal{F}_u(\cdot)$, the Fisher information of the unwarped uniform population, is approximately equal to a constant, we can rewrite this equation with the constant denoted as U:

$$\mathcal{F}_f(s) \approx U(1 + f'(s))^2. \tag{12}$$

Here, we have also rewritten the first derivative of $\hat{s}$ with respect to s following Eq (9), which we already noted is equal to the neuronal density (Eq (10)).

To consider the effect of the gain parameter on the Fisher information, we multiply each of the tuning curves in Eq (7) by a function $g(s)$, giving:

$$\mathcal{F}_g(s) = \sum_n^N \frac{[[g(s)h_n(s)]']^2}{g(s)h_n(s)}. \tag{13}$$

Then using the product rule we see,

$$\mathcal{F}_g(s) = \sum_n^N \frac{(g(s)h_n'(s) + g'(s)h_n(s))^2}{g(s)h_n(s)}. \tag{14}$$

Assuming that the gain function $g(s)$ changes slowly compared to the tuning curves, so $g(s)h'(s) \gg g'(s)h(s)$, we have:

$$\mathcal{F}_g(s) \approx \sum_n^N \frac{(g(s)h_n'(s))^2}{g(s)h_n(s)} = g(s)U. \tag{15}$$

So, the Fisher information of the warped and scaled population $\mathcal{F}_{f,g}(s)$ is:

$$\mathcal{F}_{f,g}(s) \approx g(s)(1 + f'(s))^2 U. \tag{16}$$

These approximations apply for the Fisher information of the population, rather than for any individual tuning curve. The validity of the assumptions we made to arrive at this expression for the Fisher information can be visualized by comparing the result of Eq (16) to the empirically measured Fisher information in example neuronal populations. We can see in Fig 2B and 2G that the uniform population Fisher information is approximately constant and that the example warped and scaled population Fisher information is well described by Eq (16).

Results

Optimally efficient 1-D heterogeneous neuronal populations

Using the described coding objective and neuronal population model, we will first determine the parameters of a neuronal population optimized to encode a 1-D stimulus. The constants in Eq (3) can be ignored for the purpose of optimization. So the final form of the objective

function we will use is:

$$\int p(s)\log[\mathcal{F}_{f,g}(s)]ds. \tag{17}$$

In accordance with previous work on this problem [8, 9] we adopt two constraints:

$$\int p(s)g(s)ds = R \tag{18}$$

$$\int (1 + f'(s))ds = N. \tag{19}$$

These are global resource constraints. The first is approximately equal to the expected firing rate of the population as a whole and so can be interpreted as requiring that the total number of spikes fired by the population in a given time interval will be finite and constant. Integrating the density of tuning curves yields the total number of neurons in the population, thus the second constraint expresses the fact that the encoding scheme must only use a finite number of neurons.

We are dealing with an optimization problem with integral constraints so we turn to methods from the calculus of variations. To solve such a problem, as in [8, 9], we design a function called a Lagrangian,

$$\mathcal{L}(f'(s), g(s), s, \lambda_1, \lambda_2) =$$

$$\int p(s)\log[g(1 + f'(s))^2 U]ds + \lambda_1\left(\int (1 + f'(s))ds - N\right) + \lambda_2\left(\int p(s)g(s)ds - R\right). \tag{20}$$

To optimize, we solve the system of equations obtained by setting the gradient of the Lagrangian equal to zero (note that in this case taking the partial derivative with respect to the Lagrange multipliers λ_1 and λ_2 simply reproduces the constraint equations):

$$\frac{\partial \mathcal{L}}{\partial f'(s)} = \frac{2p(s)}{1 + f'(s)} - \lambda_1 = 0 \tag{21}$$

$$\frac{\partial \mathcal{L}}{\partial g(s)} = \frac{p(s)}{g(s)} + \lambda_2 p(s) = 0. \tag{22}$$

Solving for $1 + f'(s)$ in the first equation and $g(s)$ in the second gives:

$$1 + f'(s) = \frac{2p(s)}{\lambda_1} \tag{23}$$

$$g(s) = \frac{1}{\lambda_2}. \tag{24}$$

Because $1 + f'(s)$ must integrate to N, $p(s)$ integrates to 1, and $p(s)g(s)$ must integrate to R, it is simple to determine the value of the λ's and write the solutions:

$$1 + f'(s) = Np(s) \tag{25}$$

$$g(s) = R. \tag{26}$$

Recall that $1+ f'(s)$ is equivalent to the density function from prior work, which we denote with $d(s)$ (Eq (10)) [8, 9]. Thus, substituting this identity in to Eq (25), we reach the same conclusion as prior work that

$$d(s) = Np(s). \tag{27}$$

That is, the density of tuning curves in the optimal encoding population is proportional to the probability distribution of the stimulus and the gain is constant (see Fig 1 upper row).

Because of the resulting proportionality between probability and density, and the fact that the gain function should be held constant, the optimal transformation results in a population where each neuron is equally likely to fire. In regions of high stimulus probability, there are a large number of tuning curves that are relatively narrow or selective. In regions of low probability, the tuning curves are sparser, but more broad (so they fire strongly in response to a larger range of stimuli). Said another way, the transformation to the optimal population maps the stimulus distribution to a new sensory space in which the distribution and the Fisher information are all uniform. This observation links the problem of optimal coding to the problem of finding a *mapping function* between an arbitrary probability distribution and the uniform distribution.

What is the form of the mapping function that accomplishes this goal? Prior work defined this mapping function *a priori* to be the cumulative of a function parameterizing the density of the tuning curves [8, 9, 30]. Following Eq (27), the optimal mapping function is then simply a scaled version of the cumulative of $p(s)$. However, by using the cumulative density, this formulation is limited to situations in which the transformation *can* be described by a cumulative. This limitation becomes problematic for the case of multidimensional stimuli because the cumulative treats all dimensions as if they are separable. In our new formulation, we have replaced the cumulative density with the more general transformation embodied by the variable $\hat{s}$ (that is, $s+ f(s)$). This generalization produces the same solution in 1-D, as described above. In addition, in the following section we will see that this formulation can describe the appropriate allocation of neuronal density for non-separable distributions in arbitrary dimensions in closed form.

Extending the 1-D optimization to arbitrary dimensions

For a neuronal population encoding k stimulus dimensions, the Fisher information becomes a kxk matrix with the entry in row i and column j defined by:

$$[\boldsymbol{\mathcal{F}}(\boldsymbol{s})]_{ij} = E\left[-\frac{\partial^2 \log\ p(\boldsymbol{r}|\boldsymbol{s})}{\partial s_i \partial s_j}\right]_s, \tag{28}$$

where s_i and s_j are two dimensions in the stimulus space. Note that the multidimensional stimulus is now defined by a k-dimensional vector $\boldsymbol{s}$. The k-dimensional tuning function of the nth neuron in the population will thus be written as $h_n(\boldsymbol{s})$. For the Poisson population, Eq (28) reduces to:

$$[\boldsymbol{\mathcal{F}}(\boldsymbol{s})]_{ij} = \sum_n \left(\frac{\partial h_n(\boldsymbol{s})}{\partial s_i}\right)\left(\frac{\partial h_n(\boldsymbol{s})}{\partial s_j}\right)\left(\frac{1}{h_n(\boldsymbol{s})}\right). \tag{29}$$

The objective function will now depend on the determinant of the Fisher information matrix, rather than the scalar value of the Fisher information. This is the case because in 1-D Fisher information bounds a variance, while in multiple dimensions the determinant of the Fisher information matrix bounds the determinant of a covariance matrix [11]. Thus, similar

to the 1-D population (Eq (17)), the k-dimensional population will seek to optimize

$$\int p(\boldsymbol{s})\log(|\mathcal{F}_{f,g}(\boldsymbol{s})|)d\boldsymbol{s}, \tag{30}$$

where $|\cdot|$ denotes the matrix determinant.

The expression for how the Fisher information changes under the multidimensional reparameterization must be adjusted from 1-D as well. Starting with the displacement transformation from Eq (9), which is now $\hat{\boldsymbol{s}} = \boldsymbol{s} + \boldsymbol{f}(\boldsymbol{s})$, the effect of the warping into the sensory space is:

$$\mathcal{F}_f(\boldsymbol{s}) = \boldsymbol{J}_{\hat{s}}^T \mathcal{F}_u(\hat{\boldsymbol{s}}) \boldsymbol{J}_{\hat{s}}, \tag{31}$$

where $\boldsymbol{J}_{\hat{s}}$ is the Jacobian matrix of the transformation from $\boldsymbol{s}$ to $\hat{\boldsymbol{s}}$. For square matrices, $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{A}||\boldsymbol{B}|$ and $|\boldsymbol{A}^T| = |\boldsymbol{A}|$, so we can write

$$|\mathcal{F}_f(\boldsymbol{s})| = |\mathcal{F}_u(\hat{\boldsymbol{s}})||\boldsymbol{J}_{\hat{s}}|^2. \tag{32}$$

In 1-D we made the assumption that the sum of the Fisher information as a function of the stimulus variable was approximately constant, and the corresponding assumption in multiple dimensions is that the determinant of $\mathcal{F}_u(\cdot)$ is approximately constant. We will again call this constant U.

The effect of the gain function on the Fisher information will also be different in higher dimensions. The gain function will multiply each element of the matrix, so because $|a\boldsymbol{A}| = a^k|\boldsymbol{A}|$ where k is the dimension of the matrix, we have:

$$|\mathcal{F}_g(\boldsymbol{s})| = g(\boldsymbol{s})^k U. \tag{33}$$

Finally, the determinant of the fully parameterized Fisher information matrix can be written as:

$$|\mathcal{F}_{f,g}(\boldsymbol{s})| = g(\boldsymbol{s})^k |\boldsymbol{J}_{\hat{s}}|^2 U. \tag{34}$$

After making the replacement for the density of the warped tuning curves $d(\boldsymbol{s}) = |\boldsymbol{J}_{\hat{s}}|$ (note that the determinant of the Jacobian of a transformation *defines* the relative density between the two spaces), we can write the Lagrangian for the multidimensional case (the optimization constraints are essentially unchanged from the one dimensional case):

$$\mathscr{L}(d(\boldsymbol{s}), g(\boldsymbol{s}), \boldsymbol{s}) =$$
$$\int p(\boldsymbol{s})\log[g(\boldsymbol{s})^k d(\boldsymbol{s})^2 U]d\boldsymbol{s} + \lambda_1\left(\int d(\boldsymbol{s})d\boldsymbol{s} - N\right) + \lambda_2\left(\int p(\boldsymbol{s})g(\boldsymbol{s})d\boldsymbol{s} - R\right). \tag{35}$$

Setting the gradient equal to zero yields the system of equations:

$$\frac{\partial \mathscr{L}}{\partial d(\boldsymbol{s})} = \frac{2p(\boldsymbol{s})}{d(\boldsymbol{s})} - \lambda_1 = 0 \tag{36}$$

$$\frac{\partial \mathscr{L}}{\partial g(\boldsymbol{s})} = \frac{p(\boldsymbol{s})}{g(\boldsymbol{s})} + \lambda_2 p(\boldsymbol{s}) = 0, \tag{37}$$

so we conclude that $d(\boldsymbol{s}) \propto p(\boldsymbol{s})$ and $g(\boldsymbol{s}) \propto$ a constant. This gives us a result in higher

dimensions that mirrors the 1-D case:

$$d(\boldsymbol{s}) = Np(\boldsymbol{s}) \tag{38}$$

$$g(\boldsymbol{s}) = R. \tag{39}$$

In summary, rather than parameterize a heterogenous population of tuning curves by the cumulative of the density with which they tile the stimulus space, we chose instead to use a displacement field that warps an initially uniform lattice. This choice allowed us to avoid any reference to a cumulative density function, which is beneficial because the utility of the cumulative is restricted to 1-D. However, because the objective function turns out to depend on the Jacobian of the transformation between the uniform tiling and the warped tiling, the density still appears in the k-dimensional optimization problem in a natural way. This allowed us to arrive at a solution for the optimal density with which a population of heterogeneous tuning curves should tile a multidimensional stimulus space given the probability distribution of the stimuli, and a solution for the gain function that optimally modulates the firing rate of each neuron in said population. Additionally, the result is still intuitively sensible: it says that the tiling should be more dense in regions where the stimulus is more likely, but that the response gain of tuning curves should be uniform. Taken together these statements imply that in an optimal population each neuron is equally likely to fire a spike in natural conditions.

Importantly, at this point we have not shown that a displacement field exists that can describe an optimal transformation between the stimulus space $\boldsymbol{s}$ and the sensory space $\hat{\boldsymbol{s}}$. We have rather shown that if or when this displacement field does exist, the optimal density of a population of tuning curves is proportional to the density of the multidimensional stimulus. Note also that the precise arrangement of a k-dimensional population is not uniquely determined by the density because it only describes the *determinant* of the Jacobian, which cannot be integrated to recover the mapping. This is in contrast to 1-D, where the Jacobian is simply the 1-D derivative making it possible to directly integrate the density to recover the mapping. Thus for multidimensional tuning curves, in order to show that the optimal neuronal population density is proportional to the density of the stimulus, a field $\boldsymbol{f}(\boldsymbol{s})$ that maps between stimulus and sensory space must be found such that its Jacobian determinant is properly defined. In the next section, we show that, although there is no longer a one-to-one correspondence between the density and the mapping function in higher dimensions, the mapping function is nonetheless well constrained and can be approximated with numerical methods.

Determining optimally efficient heterogeneous neuronal populations in higher dimensions

A description of the multidimensional displacement field $\boldsymbol{f}(\boldsymbol{s})$ that embodies the mapping function can be derived by considering the properties of the stimulus and sensory spaces. Following on the derivation in [27], we will show that some basic assumptions about these spaces can be used to constrain the possible transformations between them. First, we note that both the stimulus probability distribution in stimulus space $p(\boldsymbol{s})$ and the uniform probability distribution in sensory space $u(\hat{\boldsymbol{s}})$ must integrate to one. Thus, we have:

$$\int p(\boldsymbol{s})d\boldsymbol{s} = \int u(\hat{\boldsymbol{s}})d\hat{\boldsymbol{s}}, \tag{40}$$

where multiple integration over the dimensions of $\boldsymbol{s}$ and $\hat{\boldsymbol{s}}$ is implied.

To examine the relationship between the two spaces, we can eliminate the integrals by differentiating both sides of this equation with respect to $\boldsymbol{s}$. First, because $\hat{\boldsymbol{s}}$ is defined by a

transformation of s, we can let $\hat{c}(s)$ be the transformation function for sensory coordinates: $\hat{s} = \hat{c}(s) = s + f(s)$. The second integral can now be rewritten using the change of variables theorem yielding:

$$\int p(s)ds = \int u(\hat{c}(s))|J(\hat{c}(s))|ds. \tag{41}$$

Here $J(\cdot)$ denotes the Jacobian operator, and $J(\hat{c}(s))$ gives the Jacobian matrix $J_{\hat{s}}$ from Eq (31). Taking the derivative of each side with respect to the (arbitrary) upper limits of integration, substituting the definition of $\hat{c}(s)$, and isolating the determinant gives a constraint on the relationship between the two distributions, and therefore a constraint on $f(s)$:

$$p(s)/u(s) = |J(s + f(s))| = |I + J(f(s))|. \tag{42}$$

where I is the identity matrix. Here we have substituted $u(s)$ for $u(\hat{c}(s))$ because u is a constant function and the argument has no effect on its value. The limits of the multiple integration are a vector, and the derivative has been taken with respect to each element.

Note that Eq (42) is under constrained because each component of the vector $f(s) \in \mathbb{R}^k$ is an unknown, whereas the left hand side is a scalar value. However, recall from the 1-D derivation that $f(s)$ is constrained to be a smooth, continuous function of s. This implies that both the sensory and stimulus spaces, $\hat{s}$ and s, are defined over continuous coordinate systems that are simply connected and do not have coordinate singularities—an assumption that has been implicit in prior work, where smooth 1-D distributions $p(s)$ have been integrated into a cumulative density function to transform coordinates between the stimulus and sensory spaces. We will show that these mild constraints imply that $f(s)$ is a conservative field in higher dimensions, and can therefore be described as the gradient of a scalar function. This reduces the number of unknowns to one at each point s. We refer to this scalar function as the encoding potential: the gradient of this function yields the mapping that is needed to acheive the optimality criteria derived in the preceding sections.

For $f(s)$ to be conservative, its integral along an arbitraty path $r(t)$ through stimulus space must not depend upon the path taken, but solely on the endpoints. Here, t denotes the variable parameterizing the path. To write the line integral along the path, note that $f(s) = \hat{c}(s) - s$ is the difference between two coordinates. Let $c(s)$ define the stimulus coordinates similarly to the sensory coordinates. That is, $s = c(s)$, where $c(s)$ is simply the identity function. Thus, $f(s) = \hat{c}(s) - c(s)$. This additional notation is adopted to make the meaning of the line integral clear:

$$\int f(r(t)) \cdot r'(t)dt = \int [\hat{c}(r(t)) - c(r(t))] \cdot r'(t)dt. \tag{43}$$

Here, the dot is used to denote the dot product. For the line integral to be independent of the path, $r(t)$, it must be integrable. The second coordinate function, $c(s)$, integrates trivially. Let $\gamma(s)$ refer to this stimulus coordinate potential:

$$\int c(r(t)) \cdot r'(t)dt = \frac{1}{2}s^2 + C \equiv \gamma(s). \tag{44}$$

where $r'(t) = \frac{dr}{dt}$ and C is a constant of integration. Whether $f(s)$ is conservative therefore depends upon the integrability of the sensory coordinate function, $\hat{c}(s)$.

We can ensure $\hat{c}(s)$ to be Riemann integrable by leveraging a constraint that has been implicit in prior work [9]: the sensory coordinate function is constrained to be continuous and thus integrable. Note that this introduces a constraint on the coordinates, rather than the

functions defined on them. Let $\hat{\gamma}(\boldsymbol{s})$ refer to this sensory coordinate potential:

$$\int \hat{\boldsymbol{c}}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) dt \;\equiv\; \hat{\gamma}(\boldsymbol{s}). \tag{45}$$

The line integral of the displacement field $\boldsymbol{f}(\boldsymbol{s})$ can therefore be described by a scalar potential $v(\boldsymbol{s})$:

$$\int \boldsymbol{f}(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t) dt = \hat{\gamma}(\boldsymbol{s}) - \gamma(\boldsymbol{s}) \;\equiv\; v(\boldsymbol{s}) \tag{46}$$

$$\boldsymbol{f}(\boldsymbol{s}) = \nabla v(\boldsymbol{s}). \tag{47}$$

Here, $\nabla v(\boldsymbol{s})$ denotes the gradient of the scalar potential. We can now rewrite Eq (42) with the substitution $\boldsymbol{f}(\boldsymbol{s}) = \nabla v(\boldsymbol{s})$:

$$p(\boldsymbol{s})/u(\boldsymbol{s}) = |\boldsymbol{I} + \mathrm{J}(\nabla v(\boldsymbol{s}))| \tag{48}$$

$$= |\boldsymbol{I} + \mathrm{H}(v(\boldsymbol{s}))|, \tag{49}$$

where $\boldsymbol{I}$ is the identity matrix, and $\mathrm{H}(\cdot)$ denotes the Hessian operator. Note that this equation now contains only one unknown (the scalar value of $v(\boldsymbol{s})$ at each point in space), and holds regardless of the dimensionality of the spaces.

In summary, constraining the transformed sensory coordinates, $\hat{\boldsymbol{s}}$, to be continuous implies that the line integral in Eq (45) can be evaluated for all paths, which in turn implies that the underlying displacement field $\boldsymbol{f}(\boldsymbol{s})$ can be written as the gradient of a scalar potential function —and conversely, constraining $\boldsymbol{f}(\boldsymbol{s})$ to be the gradient of a scalar function implies that the coordinates $\hat{\boldsymbol{s}}$ are continuous. In the next section we show that Eq (49) can be solved for the potential $v(\boldsymbol{s})$, from which we recover $\boldsymbol{f}(\boldsymbol{s})$. The optimal neuronal population density is therefore proportional to the density of the stimulus as predicted in Eq (38).

Solving for the optimal mapping function in 2-D

The work laid out in the previous sections advances our ability to predict the properties of neuronal populations that efficiently encode more than one stimulus dimension. In practice, the characterization of neuronal tuning curves beyond 1-D requires exponentially increasing amounts of data collection in the laboratory. While large datasets of multidimensional tuning curves are becoming more feasible due to the development of high-throughput data collection paradigms, as a next step we will focus on making predictions for 2-D stimuli, because these are most likely to be relevant for experimental confirmation in the near term.

We will denote the separate dimensions of a 2-D stimulus $\boldsymbol{s}$ as s_1 and s_2. For 2-D stimuli, the determinant in Eq (49) expands to give the following differential equation:

$$p(\boldsymbol{s})/u(\boldsymbol{s}) = \frac{\partial^2 v}{\partial s_1^2}\frac{\partial^2 v}{\partial s_2^2} - \left(\frac{\partial^2 v}{\partial s_1 \partial s_2}\right)^2 + \frac{\partial^2 v}{\partial s_1^2} + \frac{\partial^2 v}{\partial s_2^2} + 1. \tag{50}$$

A numerical method for solving this equation is given in [27]. Briefly, the optimization procedure initializes $v(\boldsymbol{s})$ to the zero matrix (no transformation) at a fixed resolution. This initial estimate is then refined to minimize the total squared difference between the right and left hand sides of Eq (50), accumulated across the discretized domain (See [27], Equations (17)–(23).) Numerical refinement proceeds using the non-linear conjugate gradient method of [37] and computations are made efficient using optimal, separable filters derived by [38].

The optimization is wrapped in a multi-grid scheme in which a low resolution potential is first estimated.

The discretized domain is defined within a finite square boundary. Geometric properties of the potential are used to constrain the numerical solution within this boundary. Specifically, the potential must obey boundary constraints, or probability mass will be mapped outside of its finite domain (mass must be conserved under $f(s)$). Components of ∇v that are orthogonal to the boundary must be zero at the boundary. The potential must therefore become flat as one approaches the boundary, but can be sloped as one travels along the boundary. This constraint is enforced by repeating the values of $v(s)$ within a fixed perpendicular distance of the boundary. The fixed distance is the width of the discrete convolutional operators, guaranteeing that the perpendicular component of the discrete gradient is zero at the boundary. In the optimization, these repeated values are changed as a group to enforce this constraint.

Lastly, the above optimization as-is would force the domain boundary to be square. To allow for different boundary shapes, the uniform distribution can be defined in a smaller interior region, such as a circle. Outside of this region, both probability distributions are defined to be equal to avoid introducing discontinuities. After optimization, the gradient field is cropped to the boundary of this region. An implementation of this optimization in Matlab code is publicly available (https://github.com/eacooper/MultidimensionalEfficientCoding). Note that the numerical optimization does not guarantee that the solution is unique. Multiple minima have not, however, been observed while testing with a variety of probability distributions. Once $v(s)$ is found, its gradient gives the displacement field $f(s)$.

Summary of constraints and assumptions

Thus far, a variety of constraints, assumptions, and approximations have been chosen. Here, we will summarize and discuss some of the key choices and the limitations they introduce for predicting optimal heterogenenous neuronal populations. In the following sections, we will then apply this framework to example 2-D stimulus distributions and discuss some common features of the optimized encoding populations.

First, we make a number of choices about the initial uniform neuronal population in stimulus space. The neuronal tuning curves are assumed to be unimodal so that they can tile the stimulus space and the approximation in Eq (8) can hold. This assumption is incompatible with monotonically increasing tuning curves. For example, the population-level activity for uniformly spaced sigmoidal tuning curves increases in the direction that the sigmoids are increasing, resulting in quite different populations that satisfy the constraint on expected firing rate, as discussed in [9]. Even restricting ourselves to unimodal tuning curves, the final appearance of the optimal heterogeneous encoding population is affected by the exact shapes of these tuning curves, as well as the original density with which they tile the space. This issue is also true in prior 1-D work, in which one must manually select various features of the uniform population (the number of neurons, their initial tuning curve shapes, and their amount of overlap) in order to visualize an optimized population [9]. However, the extension into multiple dimensions necessitates additional choice-making. For example, in addition to the population density, one must also choose the lattice over which the uniform population is defined. We chose a hexagonal lattice in the simulations below, but similar results were obtained with a rectangular lattice. The initial lattice will affect the final positions of individual neurons in the optimized population—where they start affects where they end up. We assume that most biological neuronal populations are in a high density regime and should thus be evaluated at a population level, rather than based on the positioning of individual neurons. Another choice that becomes less constrained in higher dimensions is the shapes of the tuning curves. For

example, in the following simulations we chose isotropic 2-D tuning curves, but non-isotropic shapes could satisfy Eq (8) equally well. The choice of the initial tuning bandwidth in each dimension is thus left unconstrained. Importantly, these choices will change the specific details of the predicted optimal population, but they will not result in populations that disobey the core predictions of the theory: an optimal population places proportionally more neurons with narrower tuning in regions of the stimulus space that are more likely to occur and does not modify their gain.

Second, only two functions are adopted to parameterize the heterogeneous population in terms of a transformed version of this uniform population. These functions are the displacement field and the gain function. The simplicity of this parameterization limits the ultimate shape of the heterogeneous population, and also requires constraints on these functions in order to reach a closed form solution. Specifically, these functions are constrained to be continuous and slowly changing with respect to the tuning curves. These constraints ensure that Eqs (16) and (34) are valid approximations of the heterogeneous population's Fisher information, and the solution will thus hold more exactly when tuning curves are relatively narrow. These constraints are carried over from the 1-D framework that we build on, but the constraint that the displacement field be continuous takes on new significance in higher dimensions [9]. In 1-D, this continuity constraint on displacement was implicit because the *warping function* was defined by the cumulative, an integral of a density function that was itself continuous [9]. In the multidimensional case, the continuity constraint implies that the optimal displacement field must be conservative. As described in the previous section, the fact that the displacement field is conservative means that it can be written as the gradient of a scalar function and is curl free (this is the outcome of Eqs (43)–(47)). This conservative property allows us to obtain a numerical solution, but it limits our approach to modeling neuronal populations that represent continuously valued stimulus variables. For the current formulation, this constraint provides a useful starting point and importantly enables us to derive detailed predictions for populations that encode multiple stimulus dimensions. But developing models that move away from the closed-form framework used here may provide new insights into the variety of multidimensional schemes that can lead to efficient neural representations.

Lastly, the numerical method for solving for the potential function (whose gradient defines an optimal displacement field) introduces additional assumptions. Eq (50) defines the stimulus distribution and the sensory distribution (the latter being uniform) over the same domain. Because the two domains are related by the addition of a displacement field, they are required to share the same units and metric, and the left hand side of the equation is only well defined if the two domains overlap. However, the choice of domain does not change the optimal solution arbitrarily. For example, we can consider a change of units or scale in one stimulus dimension. If the range is initially 0-1 kHz but becomes 0-1000 Hz, this will change both the range of the domain in one dimension and also will generally change the geometry of the 2-D domain. Importantly, the solution to Eq (50) is invariant under such a domain transformation, after applying the same change of units to the potential function, and thus does not vary arbitrarily with a change to the geometry of the domain. In contrast, the shape of the domain does come into play when the numeric optimization is implemented. The main observed effect of the boundary shape is to change the tuning curve shapes near the boundary. We use a simple circular or square boundary in all simulations that follow. Similar issues arise in the 1-D cumulative distribution solution, in which tuning curves become constant in areas of the domain where the stimulus probability falls to zero. From a practical standpoint, because the choice of boundary is unconstrained and primarily affects the results near the boundary, we restricted our analysis to a smaller central square region of the stimulus space.

Stimulus distributions and optimal neuronal populations in 2-D

Given any 2-D stimulus probability distribution, the method described in the previous section allows us to find a displacement field with which to warp a neuronal population in order to efficiently encode the stimulus (i.e., with neuronal density proportionate to probability). In this section, we explore the properties of neuronal populations optimized with this method (Fig 3). Fig 3A illustrates four 2-D stimulus probability distributions. One distribution is Gaussian over s_1 and uniform over s_2 (top), one distribution is an isotropic bivariate Gaussian (upper middle), one distribution is an anisotropic leptokurtotic generalized Gaussian (lower middle) and one distribution is Gaussian over s_1 with a standard deviation that varies non-linearly with s_2 (bottom). For each 2-D stimulus probability distribution, we also show the inverse of the numerically optimized displacement field, which intuitively illustrates the warping of neuronal tuning curves (Fig 3B). Generating visualizations of the warped populations involves some choices, the most obvious of which is the shape of the unwarped 2-D tuning curves. For these examples, we chose bivariate isotropic Gaussians for the unwarped tuning curves, though any shape that can reasonably tile the stimulus space is possible. The resulting warped 2-D populations are shown in Fig 3C. Finally, for each population, two example 1-D samples are shown (Fig 3D). These samples show what the 1-D tuning curves in this population look like as a function of s_1 assuming s_2 was held constant at two different values. As described in the previous section, only the central region of the stimulus domain is analyzed in each panel due to limitations of the numerical method at the boundary.

Several features of the optimized populations are qualitatively notable. First, the density with which tuning curves tile a particular region of the space is proportional to the probability density in that region, as required by Eq (38) [8, 9]. Because the population is constrained to maintain a fixed overlap between neurons, the widths of the tuning curves in the population are thus heterogeneous. Qualitatively, the bandwidth appears narrower (sharper tuning curves) in areas of high probability, and broader in areas of low probability. However, the bandwidth can also vary substantially for s_1 and s_2. With respect to the maximum firing rates of the neurons, following from the fact that the optimal gain function is uniform, each neuron has the same peak firing rate associated with its 2-D tuning curve (Eq (39)). S2 Fig illustrates the measured distribution of Fisher information in each of the example populations in Fig 3C, and shows that these distributions follow the square root of the stimulus probability density as expected by Eq (34). Thus, if a population is optimal in the sense described here, it would be possible to predict the natural distribution of the stimuli being encoded from the Fisher information measured from the population responses. Such an inverse efficient coding scheme is similar to a method proposed in [39], which derives the stimulus distribution from the distribution of firing thresholds in an optimal population of spiking neurons.

Because empirical measurements of neurons often characterize a subset of their tuning properties, we next conducted simulations designed to examine how measurements of lower dimensional (1-D) tuning properties relate to the higher-dimensional (2-D) stimulus probability distribution encoded by example populations (Fig 4). These simulations can reveal insights into how to evaluate efficiency when the full dimensionality of the neuronal tuning curves (and the encoded sensory space) is unknown. Specifically, starting with each of the example neuronal populations illustrated in Fig 3C, we took a set of 1-D samples through each population, determining the 1-D tuning curves that would be obtained by holding either s_1 or s_2 constant and measuring neuronal spike rate as a function of the other variable. We then summarized each of these 1-D tuning curves by their overall gain (maximum response) and tuning sharpness (inverse of full width at half maximum response). In practice, experimenters may try to identify the tuning preference for each neuron along one dimension before characterizing
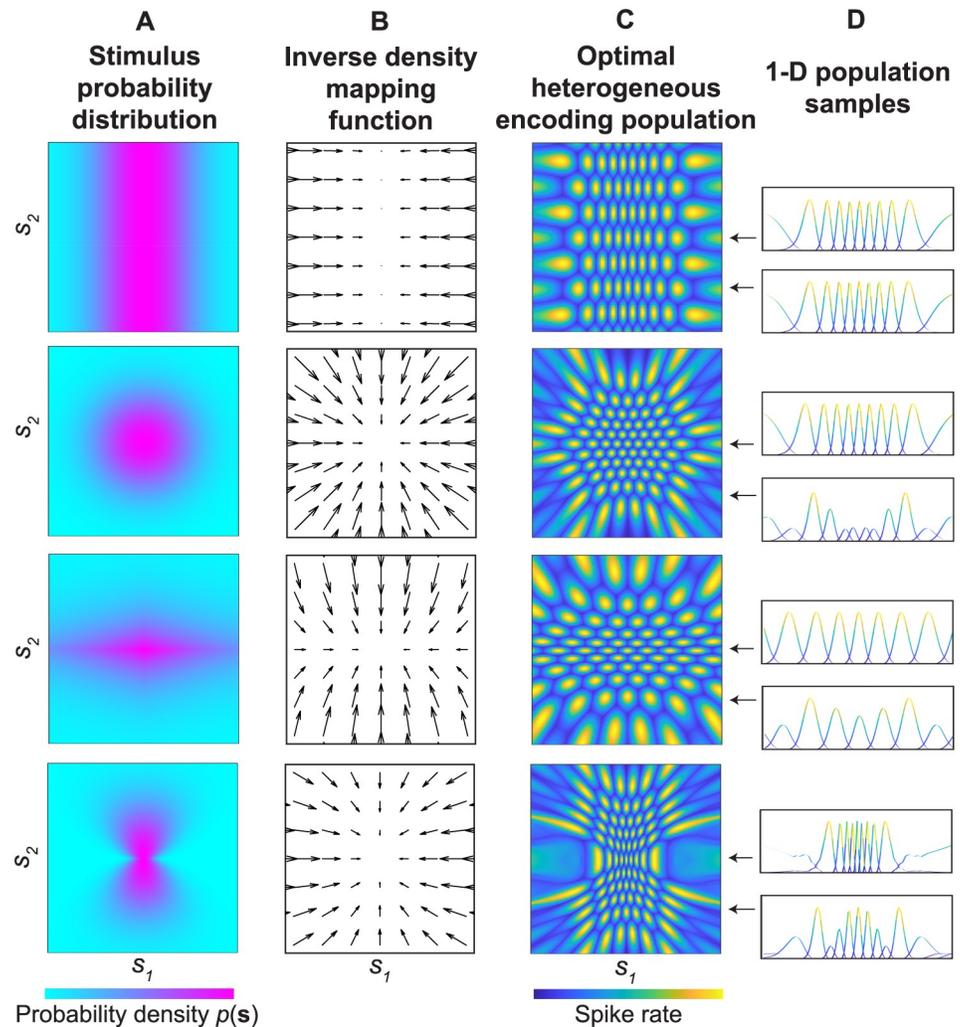
Fig 3. Example 2-D stimulus probability distributions and the resulting optimal encoding populations. A) Each row represents a different example probability distribution over two stimulus dimensions (s_1 and s_2). For each panel, the probabilities are defined over a lattice ranging from -1 to 1 (cropped to the central 65% to remove boundary artifacts). Top: uniform over s_2 and Gaussian distributed over s_1 ($\mu = 0$, $\sigma = 0.25$). Upper middle: isotropic bivariate Gaussian ($\mu = 0$, $\sigma = 0.25$ in both dimensions). Lower middle: bivariate generalized Gaussian ($\mu = 0$, $\sigma_{s1} = 0.75$, $\sigma_{s2} = 0.25$, *power* = 1.1). Bottom: Gaussian distributed over s_1, with a σ that varies non-linearly with s_2 (this distribution is non-separable). B) For each probability distribution, we show a down-sampled and scaled visualization of the inverse density mapping function. The direction and length of the arrows illustrate how density will be mapped from sensory space into the stimulus space. C) For each probability distribution, we show an example neuronal population that has been warped to optimally encode the stimulus. For these visualizations, we chose a population of neurons with isotropic bivariate Gaussian tuning curves ($\sigma = 0.05$) tiling the space on a hexagonal lattice (*spacing* ≈ 0.2). Though these choices for the population are arbitrary, varying them does not change the qualitative properties of the warped populations. Circular domain boundaries were used for the bottom three examples. To account for the uniform probability in panel A (top row), the population illustrated in panel C (top row) was defined with a square rather than a circular domain boundary. D) On the right side of each population, a pair of 1-D samples are illustrated. For each sample, s_2 is held constant and the tuning curves are visualized over s_1. Neurons with a maximum normalized response of less than 0.2 within the sample are not visualized. The distribution of Fisher information in each 2-D population is shown in S2 Fig).
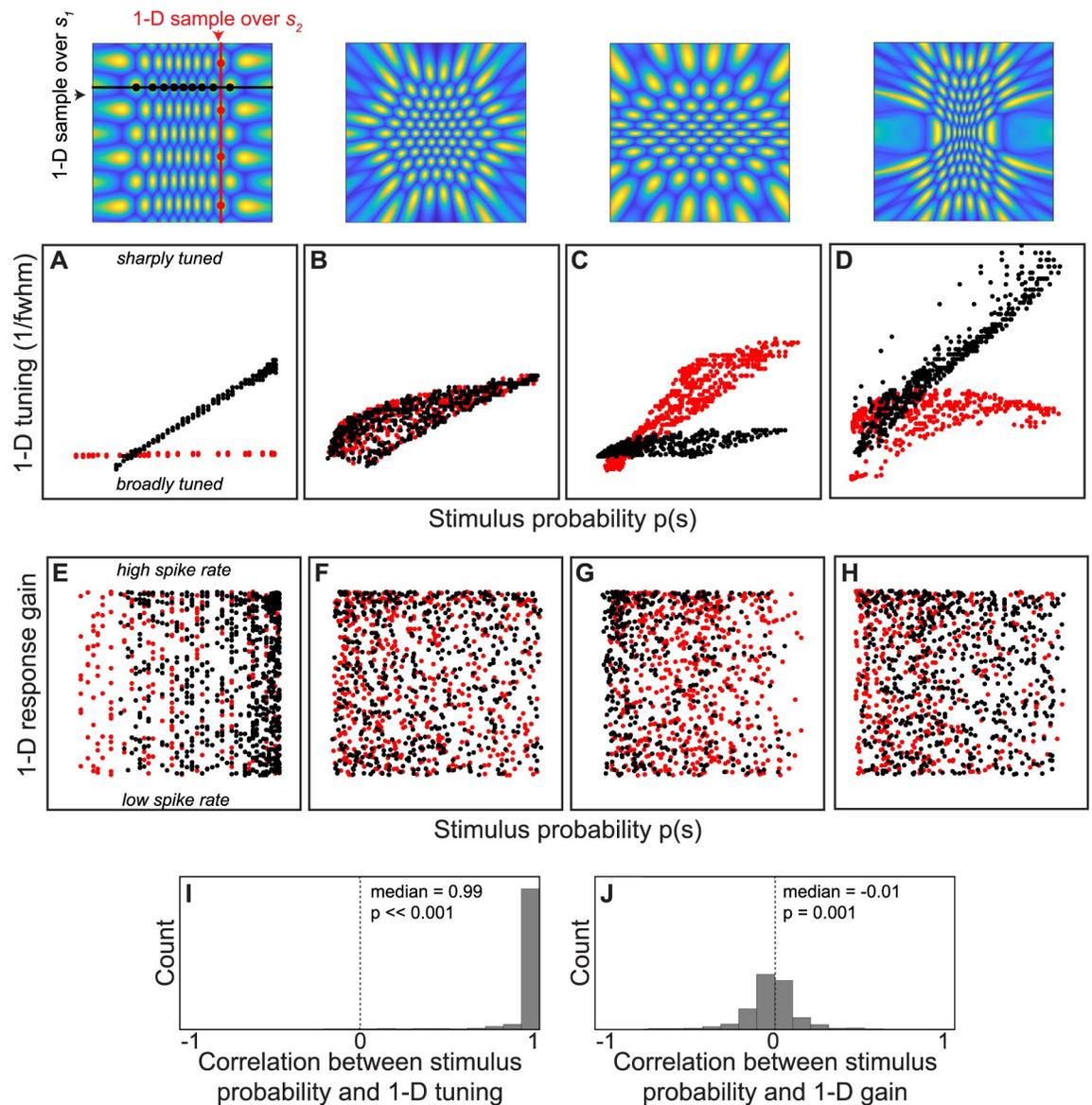
https://doi.org/10.1371/journal.pcbi.1008146.g003

Fig 4. Analysis of how lower-dimensional measurements of tuning curve properties (1-D gain and tuning width) relate to the higher-dimensional stimulus probability. Four example neuronal populations are shown, which correspond to the probability distributions and optimized mappings in Fig 3 (re-plotted in the top row). We simulated a set of 1-D experiments by selecting a single value for either s_1 or s_2 and measuring the response gain (maximum response) and tuning sharpness (inverse of the full width half maximum) of a set of neurons within this 'slice' (σ pre-warping was 0.05). This method simulates what the measured neuronal gain and tuning bandwidth would be in an experiment in which one stimulus feature was held constant and the other was varied. (A-D) For each of the illustrated populations, these panels plot the 1-D tuning sharpness as a function of probability, for a sample of neurons (400-700 neurons). Samples that were drawn by holding s_1 constant are shown in red, and samples drawn by holding s_2 constant are in black. (E-H) These panels plot the 1-D response gain as a function of stimulus probability, as in the panels above. (I,J) We repeated these simulations 500 times for randomly generated 2-D stimulus probability distributions and calculated the correlation between gain/tuning and probability. Each probability distribution was a zero-centered, bivariate Gaussian with a random orientation and major/minor σ drawn uniformly from 0.1-0.4. For each simulation, the tuning curves were modeled as isotropic Gaussians with σ drawn uniformly from 0.03-0.07. A random 1-D slice was selected, and 25 neurons were sampled. P-values indicate the results of a Wilcoxon signed rank test determining whether the median correlation was significantly different from zero.

the tuning for another (for example, identifying the preferred stimulus orientation of a visual neuron before characterizing its tuning curve for spatial frequency).

First, we asked whether measurements of 1-D tuning sharpness bear a consistent relationship to 2-D stimulus probability. Based on data from several neurophysiological studies, previous work found consistent evidence that neurons encoding more likely stimuli are more sharply tuned, however, only 1-D stimulus probabilities were considered [15]. Does the encoding of a higher dimensional stimulus space affect this prediction? In Fig 4A–4D we plot the 1-D tuning for a sample of neurons from each example population in Fig 3, as a function of the stimulus probability ($p(\mathbf{s})$, which is the joint probability of s_1 and s_2). While there is some variability, we can see that 1-D tuning generally follows a trend of having broadly tuned neurons encode low probability values, and sharply tuned neurons encode high probability values. The one exception is when s_2 is uniformly distributed and s_1 is not (Fig 4A, red circles). In this case, neurons are predicted to maintain a relatively constant bandwidth of tuning in s_2 (reflecting the uniform distribution) even when the joint probability of s_1 and s_2 is variable. However in practice it may be rare to encounter natural stimulus features that truly follow a uniform probability distribution.

In a follow up analysis, we generated 500 random stimulus probability distributions by randomly varying the aspect ratio and orientation of a bivariate Gaussian distribution. We then sampled 25 neurons from an optimal encoding population for each random distribution and calculated the tuning sharpness along a single stimulus dimension. We calculated the correlation between the stimulus probability density ($p(\mathbf{s})$) and the tuning sharpness associated with each neuron. Fig 4I shows a histogram of these correlations, which tended to be greater than zero, and typically close to one (median = 0.99). A Wilcoxon signed rank test indicated that the median was significantly greater than 0 ($p \ll 0.001$). The results of this simulation suggest that the correlation between tuning sharpness and probability is a general property across a range of stimulus distributions. Repeated runs of this simulation always produced similar results.

A separate insight arises when considering the prediction that efficient neuronal encoding populations have uniform gain. Previous work has examined the uniform gain prediction in 1-D by examining the maximum spike rate of neurons sampled from several different brain regions [15]. The authors found substantial variability of gain within populations that nominally encode the same stimulus feature, seeming to violate the uniform gain prediction. This discrepancy might suggest that gain is affected by other factors that do not relate to coding efficiency. It has also been pointed out that applying alternative tiling constraints yields a prediction that gain and probability should be positively correlated [30]. We wondered whether the observed variability in response gain might be expected from our simulation. For example, the right-most column in Fig 3 shows several 1-D samples through each 2-D population. If neuronal preferences for stimulus feature s_1 are measured empirically by varying s_1 and holding stimulus feature s_2 constant (as is often done in neurophysiological experiments), the resulting 1-D tuning curves would appear as illustrated in the insets. It is clear from these insets that the resulting data can appear as if neuronal gain is variable within this population, even though we know it is not. This apparent variability occurs simply because some neurons' peak responses to s_1 lie along different values of s_2, and vice versa. In Fig 4E–4H, we plot the 1-D response gain obtained by taking a set of 1-D samples through each example population. Similar to the empirical physiological results, we see that the predicted response gain varies between neurons, and this variability bears no consistent relationship to probability [15].

In a follow up analysis, we again varied the aspect ratio and orientation of a bivariate Gaussian probability distribution and now measured the linear correlation between probability and gain. The results suggest no consistent relationship between 1-D response gain and stimulus

probability encoded by individual neurons (Fig 4J). The median correlation value was -0.01. While this median was significantly different from zero as determined by a Wilcoxon signed rank test (p = 0.001), correlation values above 0.5 were highly unlikely. This observation suggests that in order to test predictions for any systematic patterns of gain (e.g., uniformity, correlation with probability), a more exhaustive knowledge of the stimulus features being encoded may be necessary, even if those features are statistically independent. This knowledge may be prohibitive in practice for neuronal populations that robustly encode a large number of stimulus dimensions, but could be used as a guideline for lower-level populations. At the least, future empirical work can determine gain by varying several features of a stimulus and characterizing the a multidimensional tuning curve.

Viewing 2-D populations also allows for new observations and predictions about efficient populations in higher dimensions. In addition to heterogeneity of 1-D tuning width, we see that the predicted efficient populations vary substantially in bandwidth around their peak in s_1 and s_2. For example, in each population, there are some neurons that are sharply tuned for s_1, but weakly tuned for s_2 (very broad tuning curves), and vice versa. Thus, we predict that optimal populations may contain a combination of neurons jointly tuned for each dimension, in addition to some tuned just for one dimension.

Finally, we observed that neurons in these optimal populations can exhibit heterogeneity of separability. That is to say, in many cases the efficient population contains some tuning curves that are largely separable in their tuning for s_1 and s_2 (i.e., the preferred value of one variable does not change much as the other is varied) and others that are not. Additionally, non-separable tuning curves may arise in the efficient population regardless of whether the input distribution is itself separable. For example, the bivariate Gaussian stimulus distribution in Fig 3 (upper middle row) is separable in s_1 and s_2, but produces a number of non-separable neurons (i.e., their tuning over s_1 varies as a function of s_2). The presence of both separable and non-separable curves in the efficient population is interesting when considering the problem of decoding the population responses. For sub-populations of neurons that have separable tuning curves, a single variable value can be decoded from their response without considering the value of the other stimulus dimension(s). This is because variations in the irrelevant stimuli only modulate the magnitude of each neuron's response [40]. Furthermore, if we make the observation that for many of the probability distributions we have tested the non-separable tuning curves tend to appear predominantly in regions of lower probability, and if we assume the decoding population employs a simple independent read-out method, in this scheme it appears that not only are low probability regions of the stimulus space coded with less precision (tuning curves are broader), but they also may be decoded less accurately.

Discussion

Notable theoretical advances in neuroscience have come from applying the efficient coding framework to signals derived from the natural environment (e.g., natural images, videos and sounds), which tend to be non-uniformly distributed. In several cases, this approach produces predictions about the distribution of neural resources that have been confirmed with physiological measures (e.g., [14, 15, 36, 41, 42]). Early theoretical work considered how single neurons or uniform/homogeneous populations encode sensory variables [33, 36, 43, 44]. Some of these original models explicitly considered the case of multidimensional stimuli [26]. More recent studies, motivated by the observation of substantial variability in the tuning functions of neurons in the same population, have investigated how heterogeneous populations of tuning curves can efficiently encode 1-D stimuli [8, 9, 12, 30]. However the added complexity of

multidimensional stimuli has made it difficult to generalize these heterogeneous population models to higher dimensions.

Here we have addressed this gap by providing a theoretical prediction for how neural resources should be allocated in the context of a heterogeneous population that is used to simultaneously encode an arbitrary number of stimuli. In particular, we present a closed-form solution for the optimal neuronal tiling density. This solution suggests that regions in stimulus space that are more likely should be represented by a proportionally larger number of neurons, which are each proportionally more selective. This result is a natural extension of previous work on the neural coding problem, which considered heterogeneous populations of tuning curves that respond to 1-D stimuli [8, 9]. Our results also suggest that each neuron should exhibit the same maximum response. The key intuition behind this result is that each neuron in the optimal population is equally likely to fire a spike under natural conditions (that is, when the presented stimuli come from the environmental probability distribution).

The import of any theoretical model often hinges on its ability to be verified experimentally. The prediction that neural resources simultaneously encoding multiple dimensions should be distributed proportionally to the joint environmental distribution could be experimentally verified or falsified with a combination of empirical measurements of multidimensional tuning curves and environmental statistics. Interestingly, considering the case of multidimensional stimuli revealed that the prediction that the response gain should be constant throughout such a population will likely be difficult to test experimentally. Testing that prediction would require exhaustive knowledge of all features being coded by the population, in order to rule out the possibility that observed non-uniform gain may be the result of warping along an uncontrolled stimulus dimension. Thus, we have shown that predictions made by efficient coding models can be deceptively simple, and that it is important to consider how unmodeled features of biological function (i.e., multidimensional tuning) may manifest themselves during experimental observation. In addition to these explicit predictions, this work makes it possible to ask questions about the computational benefit of utilizing an optimal multidimensional population as opposed to multiple one dimensional populations when performing downstream tasks such as decoding. For example, in [9] the density of the encoding population implicitly encodes a prior distribution during the decoding process. So the use of two 1-D populations instead of a 2-D population would encode the best separable approximation of the true joint probability distribution, which would presumably be suboptimal. The extent to which separable approximations affect decoding performance would be an interesting avenue for future research.

Like all optimization frameworks, ours makes several assumptions, fixing certain aspects of the model based on some features of biological networks and allowing others to vary during the optimization step. For example, in the current framework, the optimal mapping between input stimulus space and the sensory representation is independent of the initial pattern of tuning curves, so long as they meet the assumptions laid out in the model. It is important to note that related population coding formulations (whether they choose different features to vary or adopt different constraints) can make quite different predictions about the allocation of neural resources. This highlights that there are multiple paths to achieving coding efficiency even in the framework presented here [12, 30]. Other approaches such as sparse coding and independent components analysis address a similar class of questions about optimality in higher dimensions. Critically, our extension to multidimensional heterogenous populations allows for more direct comparisons with these and other coding schemes for higher dimensional signals. The availability of more datasets characterizing neuronal responses across multiple stimulus dimensions should eventually answer the questions of what real neuronal systems are optimized for and what constraints they operate under.

Supporting information

S1 Fig. Importance of assumptions about displacement field and gain function when parameterizing heterogeneous neuronal populations. The parameterization is illustrated as in Fig 2, however the displacement field and gain functions (C,E) now vary substantially within the bandwidth of individual neuronal tuning curves. When this is the case, the approximation of Fisher information is no longer accurate (G).
(TIF)

S2 Fig. Recovering stimulus distribution from efficient populations. For each example probability distribution from Fig 3 (A) and numerically optimized population (B), the measured Fisher information associated with the population is plotted as the determinant of the Fisher information matrix (C,D) (Eq (29)). Each panel is scaled to the maximum, which is indicated in the bottom right. The Fisher information pattern reflects the shapes and distribution of the tuning functions, which here are warped from a population of bivariate Gaussians on a hexagonal sampling lattice ($\sigma = 0.05$). In (C), neurons were relatively sparsely spaced as illustrated in (B), resulting in irregularities in the measured Fisher information ($spacing \approx 0.20$). In (D), this spacing was decreased by a factor of 2 to illustrate the smooth Fisher information. Because the determinant of the Fisher information matrix is proportionate to the squared probability of the stimulus, the results in (D) can be used to estimate the stimulus probability from the neuronal population directly (E). The panels illustrate that the numeric optimization results in a population in which the 2-D Fisher information is allocated appropriately for the input stimulus probability. In all panels, the stimulus space is cropped to +/- 0.65 to remove boundary artifacts resulting from the numeric optimization.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Thomas E. Yerxa, Eric Kee, Michael R. DeWeese, Emily A. Cooper.

Formal analysis: Thomas E. Yerxa, Eric Kee, Emily A. Cooper.

Investigation: Thomas E. Yerxa, Eric Kee, Michael R. DeWeese, Emily A. Cooper.

Methodology: Thomas E. Yerxa, Eric Kee, Michael R. DeWeese, Emily A. Cooper.

Software: Eric Kee, Emily A. Cooper.

Visualization: Thomas E. Yerxa, Eric Kee, Emily A. Cooper.

Writing – original draft: Thomas E. Yerxa, Eric Kee, Emily A. Cooper.

Writing – review & editing: Thomas E. Yerxa, Eric Kee, Michael R. DeWeese, Emily A. Cooper.

References

1. Attneave F. Some informational aspects of visual perception. Psychological review. 1954; 61(3):183. https://doi.org/10.1037/h0054663 PMID: 13167245

2. Barlow H. Possible principles underlying the transformation of sensory messages. Sensory communication. 1961; p. 217–234.

3. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381(6583):607–609. https://doi.org/10.1038/381607a0 PMID: 8637596

4. Comon P. Independent component analysis, a new concept? Signal processing. 1994; 36(3):287–314.

5. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural computation. 1995; 7(6):1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129 PMID: 7584893

6. Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annual review of neuroscience. 2001; 24(1):1193–1216. https://doi.org/10.1146/annurev.neuro.24.1.1193 PMID: 11520932

7. Geisler WS. Visual perception and the statistical properties of natural scenes. Annual review of psychology. 2008; 59:167–192. https://doi.org/10.1146/annurev.psych.58.110405.085632 PMID: 17705683

8. Ganguli D, Simoncelli EP. Implicit encoding of prior probabilities in optimal neural populations. In: Advances in neural information processing systems; 2010. p. 658–666.

9. Ganguli D, Simoncelli EP. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. Neural computation. 2014; 26(10):2103–2134. https://doi.org/10.1162/NECO_a_00638 PMID: 25058702

10. Grabska-Barwinska A, Pillow JW. Optimal prior-dependent neural population codes under shared input noise. In: Advances in neural information processing systems; 2014. p. 1880–1888.

11. Brunel N, Nadal JP. Mutual information, Fisher information, and population coding. Neural computation. 1998; 10(7):1731–1757. https://doi.org/10.1162/089976698300017115 PMID: 9744895

12. Wei XX, Stocker AA. Mutual information, Fisher information, and efficient coding. Neural computation. 2016; 28(2):305–326. https://doi.org/10.1162/NECO_a_00804 PMID: 26654209

13. Wang Z, Wei XX, Stocker AA, Lee DD. Efficient neural codes under metabolic constraints. In: Advances in neural information processing systems; 2016. p. 4619–4627.

14. Harper NS, McAlpine D. Optimal neural population coding of an auditory spatial cue. Nature. 2004; 430 (7000):682. https://doi.org/10.1038/nature02768 PMID: 15295602

15. Ganguli D, Simoncelli EP. Neural and perceptual signatures of efficient sensory coding. arXiv preprint arXiv:160300058. 2016;.

16. Cooper EA, Norcia AM. Perceived depth in natural images reflects encoding of low-level luminance statistics. Journal of neuroscience. 2014; 34(35):11761–11768. https://doi.org/10.1523/JNEUROSCI.1336-14.2014 PMID: 25164671

17. Sprague WW, Cooper EA, Tošić I, Banks MS. Stereopsis is adaptive for the natural environment. Science advances. 2015; 1(4):e1400254. https://doi.org/10.1126/sciadv.1400254 PMID: 26207262

18. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of physiology. 1962; 160(1):106–154. https://doi.org/10.1113/jphysiol.1962.sp006837 PMID: 14449617

19. Bizley JK, Walker KMM, Silverman BW, King AJ, Schnupp JWH. Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. Journal of neuroscience. 2009; 29(7):2064–2075. https://doi.org/10.1523/JNEUROSCI.4755-08.2009 PMID: 19228960

20. Yartsev MM, Ulanovsky N. Representation of three-dimensional space in the hippocampus of flying bats. Science. 2013; 340(6130):367–372. https://doi.org/10.1126/science.1235338 PMID: 23599496

21. Finkelstein A, Derdikman D, Rubin A, Foerster JN, Las L, Ulanovsky N. Three-dimensional head-direction coding in the bat brain. Nature. 2015; 517(7533):159–164. https://doi.org/10.1038/nature14031 PMID: 25470055

22. Finkelstein A, Las L, Ulanovsky N. 3-D maps and compasses in the brain. Annual Review of Neuroscience. 2016; 39:171–196. https://doi.org/10.1146/annurev-neuro-070815-013831 PMID: 27442069

23. Stemmler M, Mathis A, Herz AVM. Connecting multiple spatial scales to decode the population activity of grid cells. Science Advances. 2015; 1(11). https://doi.org/10.1126/science.1500816 PMID: 26824061

24. Born RT, Bradley DC. Structure and function of visual area MT. Annual review of neuroscience. 2005; 28:157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052 PMID: 16022593

25. Lehky SR, Kiani R, Esteky H, Tanaka K. Dimensionality of Object Representations in Monkey Inferotemporal Cortex. Neural Computation. 2014; 26(10):2135–2162. https://doi.org/10.1162/NECO_a_00648 PMID: 25058707

26. Zhang K, Sejnowski TJ. Neuronal tuning: To sharpen or broaden? Neural computation. 1999; 11(1):75–84. https://doi.org/10.1162/089976699300016809 PMID: 9950722

27. Kee E. Uniform Transformation of Non-Separable Probability Distributions. axXiv preprint arXiv:160901982. 2016;.

28. Cover TM, Thomas JA. Elements of information theory. John Wiley & Sons; 2012.

29. Abbott LF, Dayan P. The effect of correlated variability on the accuracy of a population code. Neural computation. 1999; 11(1):91–101. https://doi.org/10.1162/089976699300016827 PMID: 9950724

30. Wei XX, Stocker AA. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. Nature neuroscience. 2015; 18(10):1509–1517. https://doi.org/10.1038/nn.4105 PMID: 26343249

31. Yarrow S, Challis E, Seriès P. Fisher and Shannon Information in Finite Neural Populations. Neural computation. 2012; 24(7):1740–1780. https://doi.org/10.1162/NECO_a_00292 PMID: 22428594

32. Wang Z, Stocker AA, Lee DD. Optimal Neural Tuning Curves for Arbitrary Stimulus Distributions: Discrimax, Infomax and Minimum L_p Loss. In: Advances in neural information processing systems; 2012. p. 2168–2176.

33. Seung HS, Sompolinsky H. Simple models for reading neuronal population codes. Proceedings of the national academy of sciences. 1993; 90(22):10749–10753. https://doi.org/10.1073/pnas.90.22.10749

34. Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. Nature neuroscience. 2006; 9(5):690–696. https://doi.org/10.1038/nn1691 PMID: 16617339

35. Zemel RS, Dayan P, Pouget A. Probabilistic interpretation of population codes. Neural computation. 1998; 10(2):403–430. https://doi.org/10.1162/089976698300017818 PMID: 9472488

36. Laughlin S. A simple coding procedure enhances a neuron's information capacity. Zeitschrift für naturforschung c. 1981; 36(9-10):910–912. https://doi.org/10.1515/znc-1981-9-1040

37. Rasmussen CE, Nickisch H. Gaussian Processes for Machine Learning (GPML) Toolbox. Journal of machine learning research. 2010; 11:3011–3015.

38. Farid H, Simoncelli EP. Differentiation of discrete multidimensional signals. IEEE Transactions on image processing. 2004; 13(4):496–508. https://doi.org/10.1109/TIP.2004.823819 PMID: 15376584

39. Gjorgjieva J, Meister M, Sompolinsky H. Functional diversity among sensory neurons from efficient coding principles. PLoS computational biology. 2019; 15(11):e1007476. https://doi.org/10.1371/journal.pcbi.1007476 PMID: 31725714

40. Smolyanskaya A, Ruff DA, Born RT. Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. Journal of neurophysiology. 2013; 110(12):2806–2816. https://doi.org/10.1152/jn.00573.2013 PMID: 24089395

41. Dan Y, Atick JJ, Reid RC. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. Journal of neuroscience. 1996; 16(10):3351–3362. https://doi.org/10.1523/JNEUROSCI.16-10-03351.1996 PMID: 8627371

42. Lewicki MS. Efficient coding of natural sounds. Nature neuroscience. 2002; 5(4):356–363. https://doi.org/10.1038/nn831 PMID: 11896400

43. Snippe HP. Parameter extraction from population codes: A critical assessment. Neural computation. 1996; 8(3):511–529. https://doi.org/10.1162/neco.1996.8.3.511 PMID: 8868565

44. Pouget A, Deneve S, Ducom JC, Latham PE. Narrow versus wide tuning curves: What's best for a population code? Neural computation. 1999; 11(1):85–90. https://doi.org/10.1162/089976699300016818 PMID: 9950723