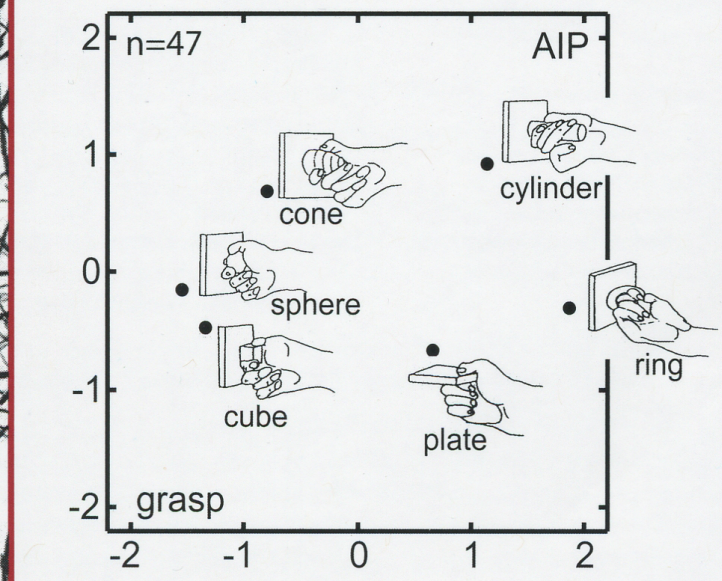
NEURAL COMPUTATION

Volume 25 Number 9 September 2013



J. Lerrive
23.VII.07

VIEW Communicated by Tim Gawne

Population Coding and the Labeling Problem: Extrinsic Versus Intrinsic Representations

Sidney R. Lehky
sidney@salk.edu
Computational Neurobiology Laboratory, Salk Institute, La Jolla, CA 92037, U.S.A.

Margaret E. Sereno
msereno@uoregon.edu
Department of Psychology, University of Oregon, Eugene, OR 97403, U.S.A.

Anne B. Sereno
Anne.B.Sereno@uth.tmc.edu
Department of Neurobiology and Anatomy, University of Texas Health Science Center, Houston, TX 77030, U.S.A.

Current population coding methods, including weighted averaging and Bayesian estimation, are based on extrinsic representations. These require that neurons be labeled with response parameters, such as tuning curve peaks or noise distributions, which are tied to some external, world-based metric scale. Firing rates alone, without this external labeling, are insufficient to represent a variable. However, the extrinsic approach does not explain how such neural labeling is implemented. A radically different and perhaps more physiological approach is based on intrinsic representations, which have access only to firing rates. Because neurons are unlabeled, intrinsic coding represents relative, rather than absolute, values of a variable. We show that intrinsic coding has representational advantages, including invariance, categorization, and discrimination, and in certain situations it may also recover absolute stimulus values.

1 Introduction

How do neurons encode the sensory, cognitive, and motor variables required to function in the world? The current consensus is that distributed representations across neural populations are central to the coding process in many situations. However, even within the population-coding paradigm, there remain questions. Here we discuss two fundamentally different ways of interpreting activities of neural populations, labeled and unlabeled, which result in extrinsic and intrinsic representations, respectively. Labeled

Neural Computation **25, 2235–2264** (2013)
doi:10.1162/NECO_a_00486

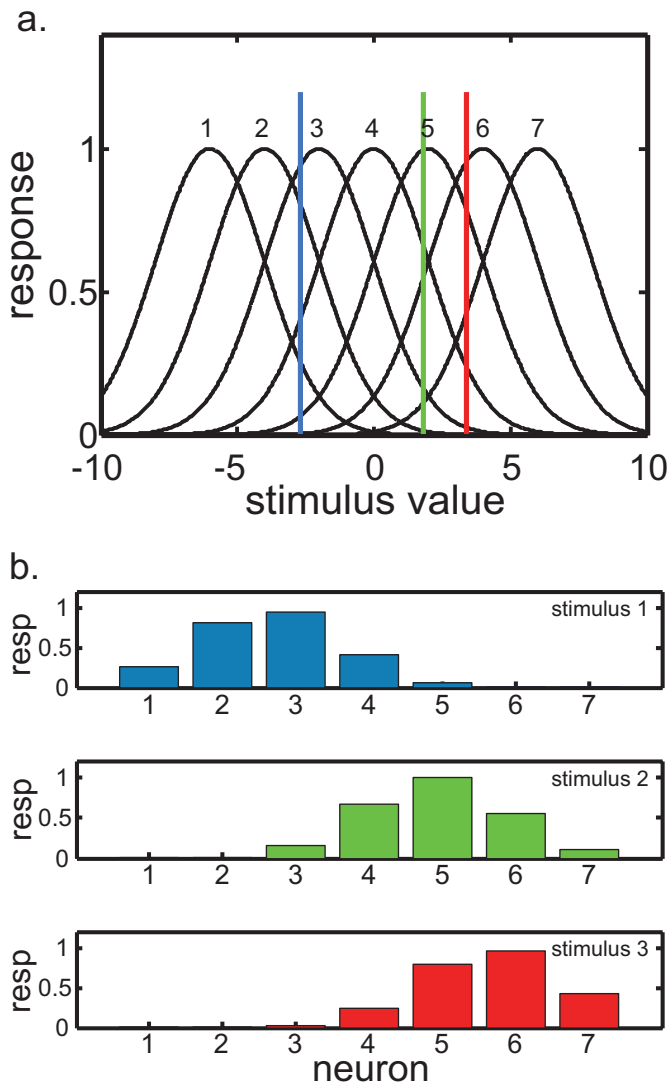


Figure 1: Schematic example of population coding. (a) Three stimulus values individually presented to a population of seven neurons with gaussian tuning curves. (b) Responses of the population to the three stimuli. (Adapted from Sereno & Lehky, 2011.)

or extrinsic representations are currently the standard approach, but unlabeled intrinsic representations are more physiological and may have inherent benefits for some aspects of stimulus representations, such as invariance and categorization.

In population coding, a variable is represented by the pattern of responses across a set of neurons (see Figure 1). Of critical importance are the relative levels of activations of different neurons. The response of each

neuron is ambiguous (more than one stimulus value can lead to the same response), but the joint activity within the population resolves that ambiguity. The paradigmatic example of population coding is a set of neurons with overlapping bell-shaped tuning curves (see Figure 1). However, tuning curves need not be bell shaped, symmetric, or even nonmonotonic (e.g., responses could be monotonic planes tilted at various angles within the parameter space). What is necessary is ambiguity in the response of each neuron, and that stimulus selectivity of different neurons overlap. (Appendix A provides historical background.)

Population activity represents a stimulus by a vector of responses $(r_1, r_2, r_3, \ldots, r_n)$, where r_i indicates the response of an individual neuron. A central concern in population coding theories is how to assign an interpretation to that response vector and extract the stimulus value. Widely used population methods include both deterministic approaches (weighted averaging of tuning curve peaks or of the tuning curves themselves) and probabilistic approaches (Bayesian estimation, maximum likelihood estimation). These extrinsic methods have been exhaustively reviewed (Averbeck, Latham, & Pouget, 2006; Földiák, 1993; Oram, Földiák, Perrett, & Sengpiel, 1998; Pouget, Dayan, & Zemel, 2000; Quian Quiroga & Panzeri, 2009; Sanger, 2003; Seung & Sompolinsky, 1993), and are summarized in appendix B. What they all have in common is that they require each neuron to be labeled with additional information (e.g., tuning curve shape, peak value, noise distribution) beyond a simple firing rate.[1] A less appreciated alternative to these extrinsic methods is intrinsic population coding, based solely on firing rate.

2 Defining Characteristics of Extrinsic and Intrinsic Population Coding

Two characteristics distinguish extrinsic and intrinsic coding: labeled versus unlabeled coding and atomistic versus relational coding.

2.1 Extrinsic Coding: Labeled, Atomistic. Extrinsic approaches to population coding require that each neuron be labeled with a parametric description of its response properties with respect to the external world. Knowing only firing rate to the current stimulus is insufficient for applying any of

[1] Nonsymmetry and other irregularities in tuning curve shape are a problem for some extrinsic methods (e.g., weighted peak averaging method) but not others (e.g., basis function or Bayesian methods). While dealing with the particulars of tuning curve shape is not a conceptual problem for extrinsic methods in general, it does present a practical data collection problem during the application of many such methods. That is due to the difficulty in measuring the detailed shapes of tuning curves (or probabilistic functions dependent on tuning curve shape) for an entire population in order to label neurons with that information.

these algorithms. For example, according to Pouget et al. (2000), maximum likelihood estimation requires "precise measurement of the tuning curves and noise distributions of each neuron." The particular extrinsic method used determines which label is required. Weighted peak averaging, for example, requires that neurons be labeled only with the values of their tuning curve peaks.

Labels within extrinsic coding provide a coordinate system or reference frame, independent of the neural firings themselves, that allows extraction of absolute stimulus values from the population response. Since labeling is in terms of some external (nonneural) variable tied to a state of the physical world, it provides external information that allows the population to be interpreted in physical world coordinates. These labels associated with extrinsic coding provide an external frame of reference that allows individual stimuli to be represented independent of other stimuli. Extrinsic coding therefore represents stimuli in an atomistic manner.

2.2 Intrinsic Coding: Unlabeled, Relational. In contrast, intrinsic coding bases its representation purely on neural firing rates without any additional information. Intrinsic coding, having no labels to the external world, cannot represent the absolute value of a single stimulus, only relative values of multiple stimuli. It therefore represents stimuli in a relational manner. Many current implementations of this approach use multidimensional scaling methods (we provide examples later in this view). Advocacy for the potential usefulness of representing stimuli relationally rather than atomistically can be found in experimental psychology (Shepard & Chipman, 1970) (calling the concept *second-order isomorphism*), as well as computational vision (Edelman, 1998, 1999) and philosophy of mind (Churchland, 2012; Churchland & Churchland, 2002). The distinction between relational and atomistic representations in population coding has antecedents in different philosophical traditions on the nature of representation (see appendix C).

3 Extrinsic Coding and the Labeling Problem

How is the labeling of neural activity required by extrinsic population coding implemented physiologically? How is this precise labeling transmitted at each synapse? These questions that follow from the labeling hypothesis, perhaps in some cases odd seeming, have not been explicitly recognized in the extrinsic coding literature, much less addressed. In the laboratory, we can externally label neurons through preliminary calibration experiments and then use the results for later population decoding. Figure 2 illustrates this labeling process for a population of motor neurons controlling the direction of arm movements, using weighted peak averaging as the decoding method (Georgopoulos, Caminiti, Kalaska, & Massey, 1983; Georgopoulos, Kalaska, Caminiti, & Massey, 1982). For each neuron, responses were

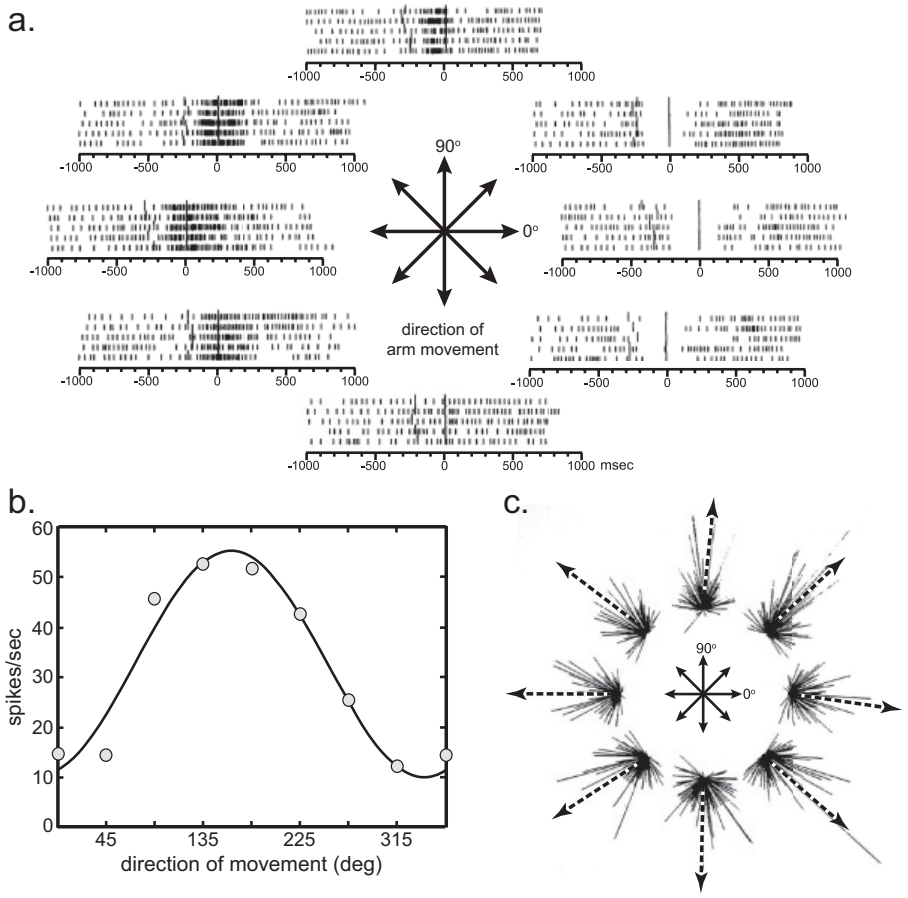


Figure 2: Labeling neurons for extrinsic population coding. (a) Activity of example neuron in motor cortex when monkey performed two-dimensional arm movements in different directions. (b) Tuning curve for direction of arm movement for the example neuron. (c) Interpretation of population activity in motor neurons using weighted averaging of tuning curve peaks. Black lines indicate responses of individual neurons. Line length is a function of firing rate, and line orientation indicates the direction of tuning curve peak. Orientation of dashed black lines shows arm movement directions assigned to population activity, calculated by peak averaging of individual neural responses. For this procedure to work, each neuron must be labeled with the value of its tuning curve peak, derived from the sort of data shown in panels a and b. (Panels a and b adapted from Georgopoulos et al., 1982; panel c adapted from Georgopoulos et al., 1983.)

collected for different movement directions (see Figure 2a), and a tuning curve was fitted to that data (see Figure 2b). To perform the population decoding (see Figure 2c), each neuron was then labeled with the value of its tuning curve peak, which provided necessary external information in addition to its firing rate. If we are interested in understanding how population coding operates in vivo (i.e., during normal brain processing) using

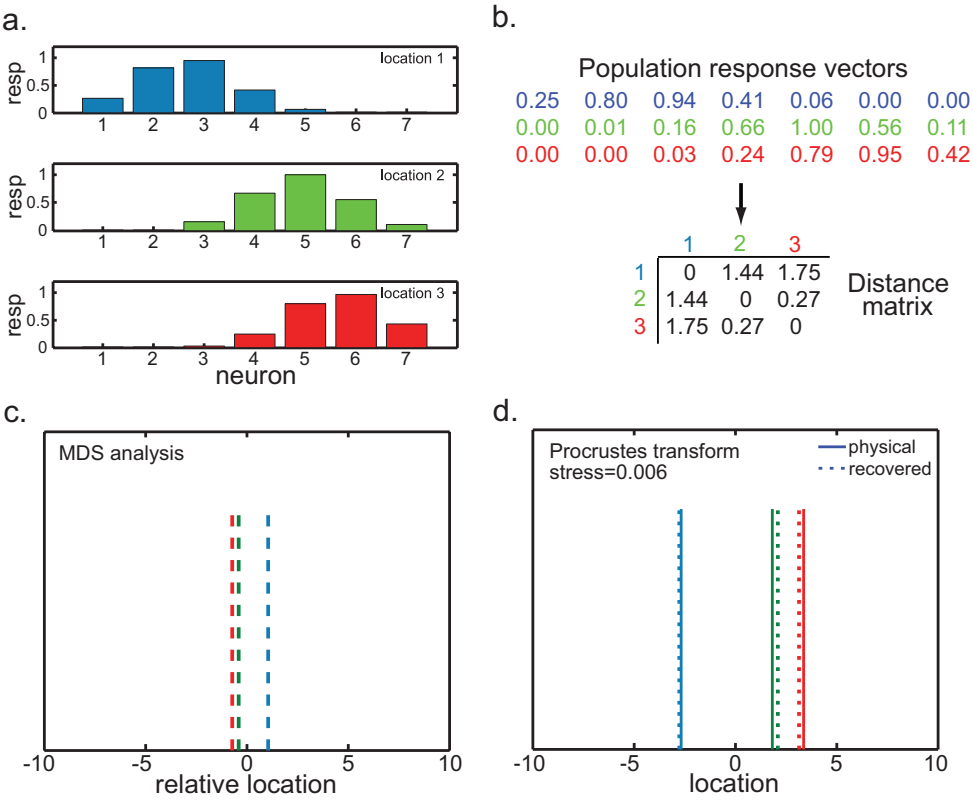


Figure 3: Intrinsic population decoding: Multidimensional scaling (MDS). Example based on the tuning curves and stimuli in Figure 1a. (a) Responses of neurons in the population to the three stimuli. (b) Response vectors for three stimuli, which are the histograms in panel a displayed numerically. Distances between response vectors are displayed in a distance matrix, which serves as immediate input to the MDS algorithm. (c) Output of the MDS algorithm, providing relative values of the three stimuli. (d) Procrustes transform of the MDS output. Solid lines are original stimulus values as shown in Figure 1a, and dashed lines are Procrustes-transformed MDS output. Stress is an error measure. (Adapted from Sereno & Lehky, 2011.)

extrinsic methods, then we cannot ignore the problem of defining the biological basis of neural labeling inherent in such methods.

4 Multidimensional Scaling: An Illustration of Intrinsic Representation

Multidimensional scaling (MDS) (Borg & Groenen, 2010; Shepard, 1980) is an example of an intrinsic approach to interpreting population activity. It utilizes only firing rate and does not require additional labeling. An illustration of MDS is given in Figure 3, based on the same population of seven neurons and the same set of three stimuli portrayed in Figure 1.

Population responses to the three stimuli are shown in the histograms of Figure 3a. Presenting those histograms in numerical form (see Figure 3b) gives three response vectors. Crucially, no additional information (e.g., labeling of tuning curve peaks) is given in MDS to these individual response vectors, unlike what occurs with labeled coding in extrinsic representations. Rather, with intrinsic representations, we are interested in the relative values (differences) between these response vectors (relational coding). Therefore, distances between all response vectors are calculated, producing a distance matrix (see Figure 3b). (The distance metric used is $d = 1 - r$, where r is the correlation between pairs of response vectors.)

The distance matrix serves as input to the MDS algorithm, which performs a dimensionality reduction (see appendix D) on the original seven-dimensional population representation. MDS is able to reduce the three points in the original 7D space to three points in a 1D space while still keeping distances between points (relative positions) almost identical (see Figure 3c). This implies that population responses to different stimuli are confined to a 1D subspace embedded within the 7D representation space, mirroring the 1D nature of the stimulus variable. (Figure 11 illustrates dimensionality reduction for a 3D representation space.) Because we cannot visualize high-dimensional spaces, we cannot easily see that relationship among population responses to different stimuli. By doing dimensionality reduction, MDS makes that structure apparent.

Because we are dealing with the population in an intrinsic manner, at no point were labels attached to neurons. Only firing rates were used. The MDS output (see Figure 3c), based on unlabeled neural activity, recovers only relative stimulus values. Compared to the original values (see Figure 1), the scale is different and the stimulus order reversed, but relative values are quite accurate. Precisely how accurate can be quantified using a Procrustes transform (see Figure 3d).

Although an intrinsic representation was derived here from comparisons across multiple sensory stimuli, in principle one could also have comparisons between current sensory inputs and memory traces of previous sensory inputs (Shepard & Podgorny, 1978). As Edelman (1999) suggested, such memory traces could act as reference landmarks or prototypes within a neural representation space, against which incoming stimuli could be relationally encoded.

MDS or other dimensionality-reduction methods (see appendix D) do not cause responses to lie on a low-dimensional manifold (i.e., subspace) within the high-dimensional neural representation space. Rather, they report whether such a low-dimensional manifold exists. We (Sereno & Lehky, 2011) and others (Churchland, 2012; Edelman & Intrator, 1997; Seung & Lee, 2000) have previously suggested that such low-dimensional representations may be computationally advantageous in some cases, for example, to more efficiently interface or communicate with other cortical areas. These low-dimensional representations need not be made explicit but could

remain implicit as low-dimensional subspaces embedded within a high-dimensional space defined by the size of the encoding neural population.

Dimensionality reduction itself is not necessary for intrinsic representations. Despite the potential advantages of low-dimensional encoding, the existence and usefulness of intrinsic encoding is conceptually independent of whether representations are low dimensional or high dimensional.

5 Intrinsic Coding: Categorization and Discrimination

Categorization and discrimination, often considered separately, both involve relationships among different stimuli (Lehky & Sereno, 2007). Stimuli within the same category would be expected to cluster in the same region of a representation space. Stimuli outside the category are more distant, perhaps in another cluster. Discrimination suggests that even if the stimuli are within the same category or cluster, they are far enough apart to be reliably distinguished. Relational representations provided by intrinsic coding make these geometrical relationships within the representation space explicit.

In contrast, the atomistic representations that are produced by extrinsic coding provide no inherent basis for clustering and discriminating stimuli. To do categorization using an extrinsic representation, for example, would require an additional level of processing to make explicit the geometrical relationships among different stimuli within the parameter space. Intrinsic representations already have such geometrical relationships built in as an inherent part of the decoding process. For clarification, if clustering or discrimination algorithms are applied to labeled neurons but the labeling information is not used in the algorithm, then this would be intrinsic, not extrinsic, coding.

To examine how population responses to different stimuli cluster, a number of studies have applied MDS to data from monkey cortex (Kayaert, Biederman, & Vogels, 2005; Kiani, Esteky, Mirpour, & Tanaka, 2007; Lehky & Sereno, 2007; Murata, Gallese, Luppino, Kaseda, & Sakata, 2000; Op de Beeck, Wagemans, & Vogels, 2001; Rolls & Tovée, 1995; Young & Yamane, 1992). Here we highlight three examples: (1) visual responses in anterior inferotemporal cortex (AIT) to simple 2D geometric shapes (Lehky & Sereno, 2007), (2) visual responses in AIT to faces (Young & Yamane, 1992), and (3) nonvisual activity in anterior intraparietal cortex (AIP) associated with different hand-grip shapes while grasping 3D objects (Murata et al., 2000).

In these studies, the population response to each stimulus was a point within an n-dimensional representation space where n was sample population size. To visualize relationships among population responses to different stimuli, those high-dimensional representations were reduced to two dimensions using MDS, keeping distances between stimuli as unchanged as possible. The results show relative positions of shapes within the neural representation space (see Figure 4), with a clustering of conditions with

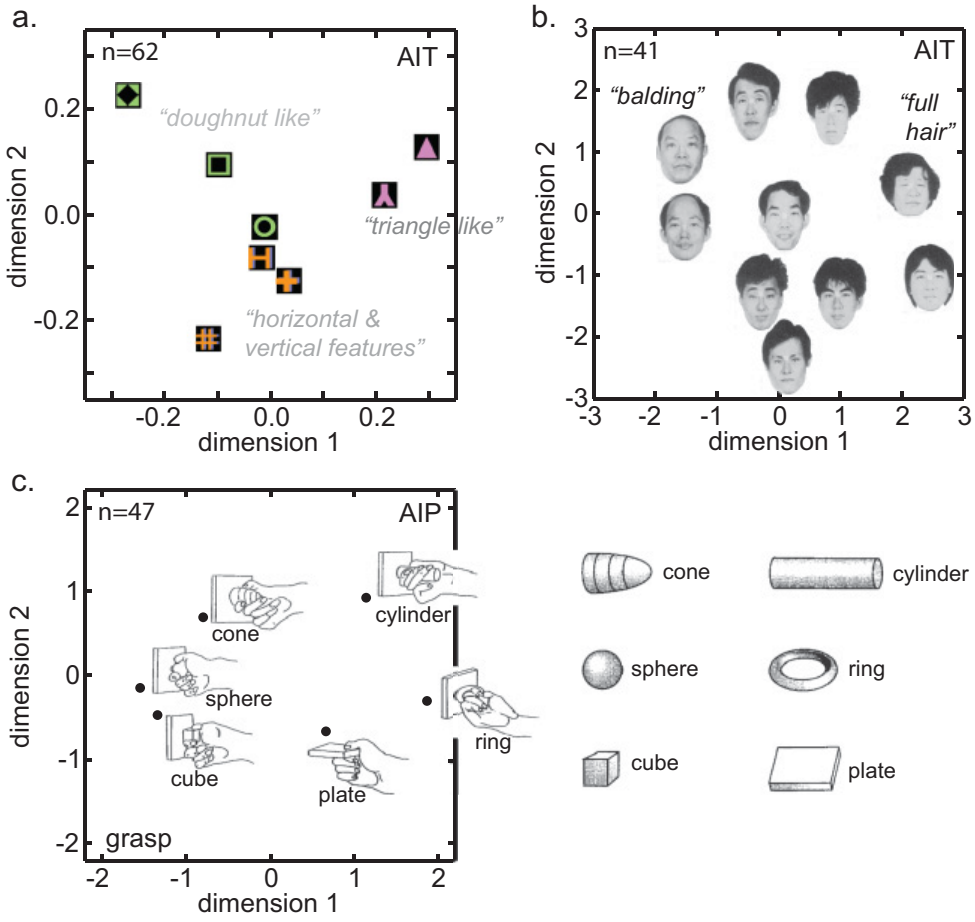


Figure 4: Examples of intrinsic population coding, based on multidimensional scaling. (a) Relative positions of neural responses to simple geometrical shapes, within a shape space derived from a population of cells in anterior inferotemporal cortex (AIT). (b) Relative positions of neural responses to faces, within a face space derived from a population of AIT cells. (c) Relative positions of neural activities corresponding to different hand grips when grasping objects having various shapes, within a hand-shape space derived from a population of cells in the anterior intraparietal (AIP) area. (Panel a adapted from Lehky & Sereno, 2007; panel b adapted from Young & Yamane, 1992; panel c adapted from Murata et al., 2000.)

similar characteristics. Likewise, data based on much larger sets of visual shapes than in these examples have yielded impressive demonstrations of categorization when interpreting population activities using intrinsic methods (Kiani et al., 2007; Kriegeskorte et al., 2008), reviewed by Kriegeskorte (2009).

This approach to population coding stands in contrast to decades of physiological work struggling to understand the features and categories

of neural representations underlying object recognition through detailed characterization of individual inferotemporal neurons, including attempts to identify their optimal stimuli (essentially tuning curve peak) (Freiwald, Tsao, & Livingstone, 2009; Fujita, Tanaka, Ito, & Cheng, 1992; Logothetis, Pauls, & Poggio, 1995; Richmond, Optican, Podell, & Spitzer, 1987; Tanaka, Saito, Fukada, & Moriya, 1991; Yamane, Carlson, Bowman, Wang, & Connor, 2008), reviewed by (Kourtzi & Connor, 2011; Logothetis & Sheinberg, 1996; Tanaka, 1996). Such information is required to implement an extrinsic population code. Intrinsic methods, using data-driven (agnostic) techniques such as MDS (see Figure 4), are able to reveal relationships inherent in the responses of neural populations to different object stimuli without any a priori knowledge or assumptions about the properties of individual neurons or the structure of the categorization.

6 Intrinsic Coding: Representation of Visual Space

A number of neurophysiological studies have used MDS to analyze population coding of visual shape (Kayaert et al., 2005; Kiani et al., 2007; Lehky & Sereno, 2007; Murata et al., 2000; Op de Beeck et al., 2001; Rolls & Tovée, 1995; Young & Yamane, 1992). Only one has applied an intrinsic approach to visual space (Sereno & Lehky, 2011). Using unlabeled neurons, it produced a representation of visual space that was relational rather than atomistic. Therefore, the global structure of space came out naturally without additional assumptions or analyses. It was possible to extract relative stimulus positions from neural populations not only in the dorsal visual stream (lateral intraparietal cortex) but also the ventral stream (anterior temporal cortex) of monkeys. Further, whereas the dorsal representation of space was quite metrically accurate, the ventral stream representation was only topologically (or categorically) accurate.

A widespread view in studies of monkey extrastriate visual processing is that large RFs throw away spatial information to produce spatially invariant object representations by pooling spatially localized responses received from earlier levels (e.g., Tanaka, 1996; Gochin, 1996, in the neurophysiology literature; Riesenhuber & Poggio (1999) in the theoretical literature). Instead, in a modeling study using intrinsic coding, large RF diameters produced the most accurate reconstructions of space (Lehky & Sereno, 2011). The better performance of large RFs in intrinsic coding holds true whether the population is noise free (see Figure 5) or noisy (see Figures 6a and 6b). In contrast, small RF diameters, as would occur in the earliest visual areas, produced poor representations of space (see Figure 5c).

7 Extrinsic Versus Intrinsic Spatial Coding

Modeling shows that optimal receptive field (RF) characteristics for coding visual space are strikingly different depending on whether extrinsic or intrinsic population coding is used. Figure 6 directly compares extrinsic

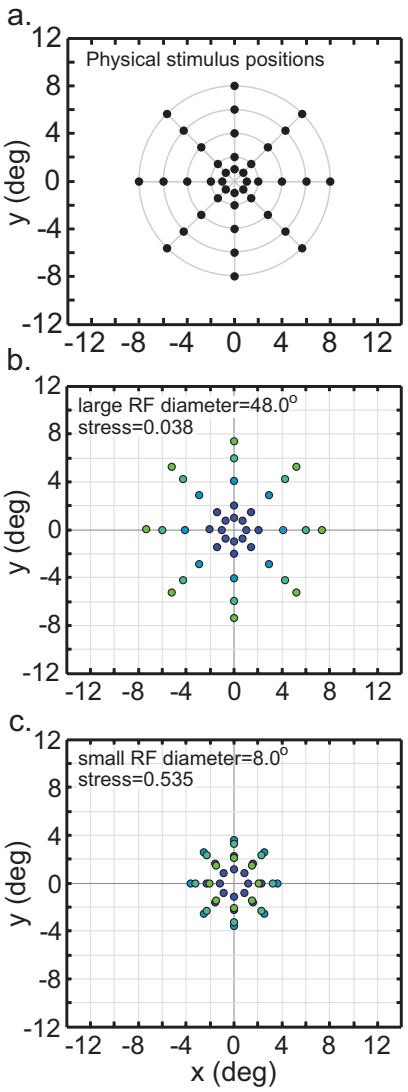


Figure 5: Example of intrinsic coding of visual space. Multidimensional scaling was used to recover stimulus locations from a population of model neurons without noise. The radius of receptive fields was defined as one space constant σ of gaussian tuning curve, so the diameter was 2σ. Spacing between RF peaks was 0.25σ, although previous work (Lehky & Sereno, 2011) shows that results are independent of RF spacing for noise-free systems. (a) Physical stimulus locations. Forty locations are arranged in a radial grid. (b) Recovered positions using large receptive fields, producing an accurate representation of space. (c) Recovered positions using small receptive fields, producing a highly distorted representation of space. Locations in the outer ring (lightest green) have curved inward, so that the representation is not even topologically accurate. In panels b and c, recovered positions were linearly rescaled by a Procrustes transform to allow quantitative comparison with physical locations. Stress is an error measure, with smaller values indicating better fit between recovered locations and physical locations. (Adapted from Lehky & Sereno, 2011.)

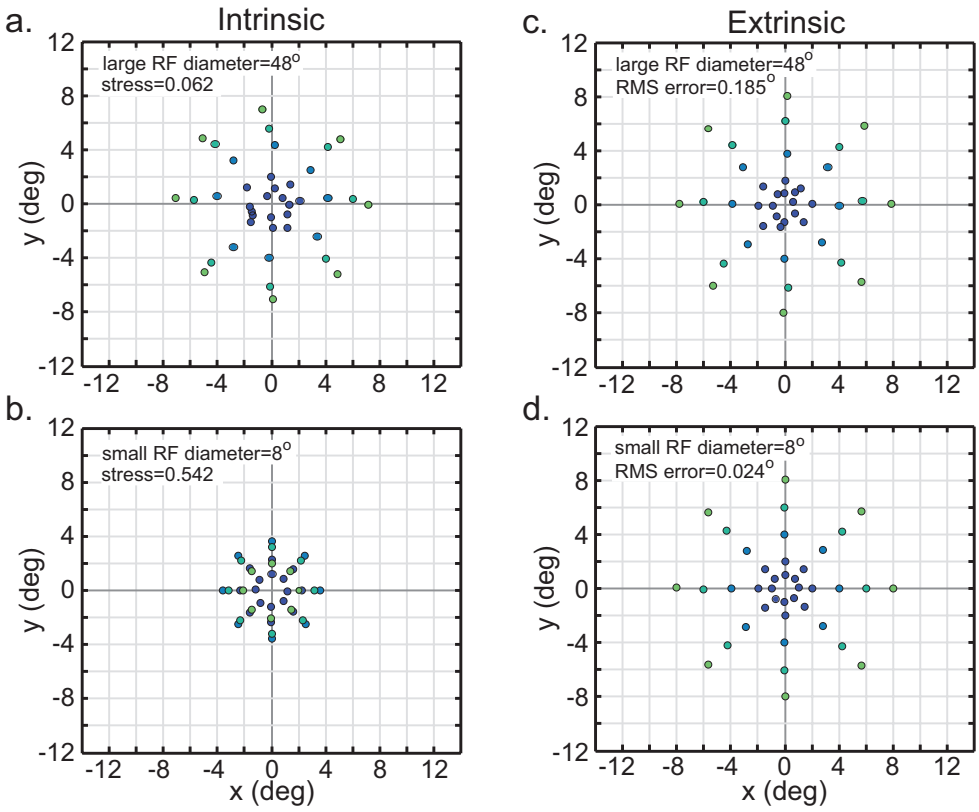


Figure 6: Comparison of population coding of visual space under intrinsic and extrinsic methods, using noisy neural populations. Population characteristics were identical in each case. The radius of receptive fields was defined as one space constant σ of gaussian tuning curve, so the diameter was 2σ. Spacing between RF peaks was 0.25σ. Uncorrelated gaussian noise was proportional to neural responses, with a standard deviation of noise equal to 0.3 of response amplitude for each neuron. (a, b) Intrinsic coding, using multidimensional scaling on unlabeled neurons with large and small RFs. Details as in Figure 5. Performance was better with large RFs. (c, d) Extrinsic coding, using weighted peak averaging on labeled neurons with large and small RFs. Performance was better with small RFs.

(weighted peak averaging) and intrinsic (MDS) methods using identical noisy populations. As we have already seen, large RFs are best for intrinsic coding. The opposite occurs using extrinsic coding. With extrinsic coding, when each neuron is labeled with the spatial location of its RF, then small RFs can produce more accurate representations of space than large RFs.

The improved accuracy of the extrinsic coding with small RFs shown in Figure 6 depends on the assumption that relative RF overlap remains constant, independent of RF size. Two consequences of keeping RF overlap constant are the following. First, the spacing between RF peaks becomes

smaller for small RFs and larger for large RFs, with performance degrading as RF overlap decreases (RF spacing increases). Second, the number of neurons in the population changes for different RF sizes to cover a given region of visual space, with performance likely degrading as number of neurons decreases. Thus, the improved performance for small receptive fields using extrinsic coding given constant RF overlap may be due to smaller spacing between neurons and greater number of neurons.

We believe that the assumption of constant RF overlap, which underlies the demonstration that extrinsic coding performs better with small RFs, is reasonable, particularly in light of studies of overlap in retinal ganglion cells (Borghuis, Ratliff, Smith, Sterling, & Balasubramanian, 2008; DeVries & Baylor, 1997), though see Zhang and Sejnowski, 1999, for an alternative approach using constant population size. By adapting the analysis of Zhang and Sejnowski to the condition of constant relative RF overlap, the improvement for small receptive fields in extrinsic coding can be seen mathematically.[2] However, correlated noise may place limits on the benefits of very small RF diameters for extrinsic coding; see (Pouget, Deneve, Ducom, & Latham, 1999).

A further difference between extrinsic and intrinsic population coding involves the minimum population size required to encode stimuli. For extrinsic coding, merely three such labeled overlapping RFs are in principle able to define stimulus position in two dimensions by a process of trilateration (or more generally a minimum of D + 1 receptive fields for a D-dimensional parameter space), whereas intrinsic coding requires a larger population to give reasonable results. For instance, for the one-dimensional (D = 1) example of intrinsic coding in Figure 3, if the number of tuning curves were reduced from seven to two, the decoding process could not recover the correct topological order of the three stimuli despite being in a noise-free system.

Thus, extrinsic and intrinsic approaches have quite different properties regarding population encoding of space and come to distinct and divergent understandings as to the role of large RF diameters in the reconstruction of accurate representations of space. Neuropsychology indicates that various higher-level cortical areas with large RFs are important in spatial representations (Jeannerod & Jacob, 2005), suggestive of intrinsic coding playing a role.

[2] Fisher information J for a population, which describes encoding accuracy, is given by $J = \eta\sigma^{D-2}K_{\phi}(F, \tau, D)$, where η is the density of RFs covering the parameter space, σ defines tuning width, D is the dimensionality of the receptive fields, and K is a function describing RF properties and stimulus duration. With a constant relative overlap (e.g., tuning curve peaks separated by 2σ), the density of receptive fields is inversely proportional to RF diameter, $\eta \propto 1/\sigma^{D}$. That makes Fisher information inversely related to receptive field size: $J \propto 1/\sigma^{2}$ (with dependence on D disappearing). In other words, encoding accuracy increases for small receptive fields.

8 Intrinsic Coding: Invariance

Representing relative values rather than absolute values using intrinsic coding has benefits for creating invariant representations. Indeed, maintaining relative values unchanged in the face of various transforms captures the essence of what an invariant representation is. As an example, if the relative positions of object features are encoded, those relationships remain unchanged if the object is translated or scaled. In this case, a relational representation can simplify the extraction of invariances compared to an atomistic representation. Within an intrinsic coding framework, invariance becomes a population property rather than a property realized in the responses of individual cells.

The contribution of population properties to spatially invariant representations using intrinsic coding has been further discussed in Lehky and Sereno (2011) and Sereno and Lehky (2011). Translational invariance was found to be sensitive to receptive field dispersion (i.e., the visual field range over which receptive field centers extend), a population characteristic that varies across visual cortical areas. Also, it is not necessary for individual neurons to be scale invariant in order for the population as a whole to be scale invariant under intrinsic coding. If neurons in the population were homogeneous with respect to their individual sensitivities to scale (have the same scale sensitivity), a much weaker condition, then under intrinsic coding, the population as a whole, would remain scale invariant. Moving to the opposite extreme of a completely inhomogeneous population, simulations (Lehky & Sereno, 2011) indicate that if responses of individual neurons are perturbed randomly in a population (equivalent to random scale sensitivity), variations average out and relational encoding within an intrinsic framework is minimally affected.

9 Intrinsic Coding and the Grounding Problem

Having access only to relative stimulus values works fine for some situations. For example, noting if a window is opened or closed can be done with relational coding of positions. However, relational coding can lead to apparent problems when physically interacting with the world, as in visual control of motor actions (grasping the window to close it).

Unlabeled neural activities underlying intrinsic, relational representations have no real-world scales associated with them, such as degrees of visual angle. Without attaching relational representations to an external scale, the representations are not grounded to the world. We call this the grounding problem (Harnad, 1990; Searle, 1980). The grounding problem does not exist for extrinsic coding, as neural labels provide an external scale for the activity of each neuron.

One way to solve the grounding problem is by allowing different relationally encoded variables (e.g., sensory and motor) to become consistent

with each other and the world (i.e., grounded to the world) through interactions with the world. This would involve experience-dependent learning during sensory-guided motor actions (Krakauer, Pine, Ghilardi, & Ghez, 2000; Salinas & Abbott, 1995; Wallman & Fuchs, 1998) (see also Churchland, 2012). For example, to make a saccade to a target, it is not necessary to indicate target location using visual cells whose spatial tuning curves are labeled in degrees of visual angle or to produce the saccade with motor cells labeled in degrees. The population coding of both can each use intrinsic, relational scales with arbitrary relationships to the physical world. As long as the intrinsic representations of perception and action are consistent and produce useful behavior in the world, the system is calibrated to the world and these intrinsic population representations are grounded.

10 Discussion

Extrinsic approaches to population coding require that all neurons be externally labeled. How such labeling is implemented (if ever) and where it occurs in the neural circuitry is unknown. Extrinsic descriptions of population coding therefore remain incomplete from a biological perspective, and perhaps even unphysiological. Unfortunately the labeling problem has received little attention within neurophysiological theories of population coding.

Intrinsic representations provide an alternative approach that sidesteps the whole labeling problem. We have delineated critical and consequential differences among the two classes of population models. In addition, we have suggested situations where intrinsic coding may be superior (e.g., categorization at the population level, representational invariances), as well as presented experimental successes for the intrinsic coding approach in neurophysiology. We propose that much neural processing uses unlabeled neurons, leading to intrinsic representations.

It is possible that extrinsic representations also exist in the brain, but this would require finding and resolving the physiological basis of whatever neural labeling is presupposed by the particular extrinsic approach. Although intrinsic and extrinsic methods are two fundamentally different approaches to population coding, it is possible that a mixture of intrinsic and extrinsic methods might be appropriate to attack a given problem.

Intrinsic population representations may also have potential for applications other than neurophysiology. Analogous procedures can be used to interpret responses from populations of voxels in fMRI (Kriegeskorte et al., 2008). This method may also be useful for neural modeling, for example, to interpret hidden layer activity or the activity of the output of supervised or unsupervised learning models, even in cases where the input or output layers are trained using extrinsic labeling. Understanding representations learned in the deeper layers of multilayer networks has been highlighted as an important issue for future research in artificial neural networks (Hinton,

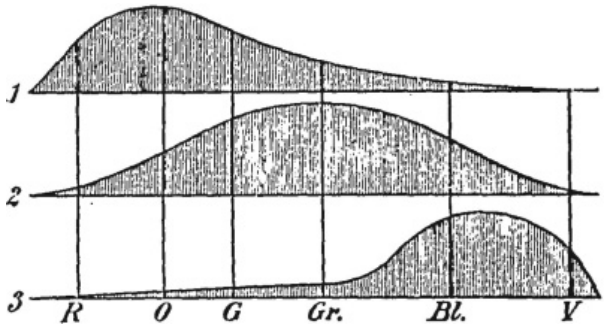


Figure 7: The first population coding model, showing three color tuning curves. These schematic curves were created by Helmholtz (1909/1962) based on an idea by Young (1802), and were first published in 1860.

2007). Furthermore, intrinsic population representations can have practical clinical applications. For example, current approaches to brain-machine interfaces (BMIs) generally require labeling in order to interpret population activities (Bokil, Pesaran, Andersen, & Mitra, 2006; Gao, Black, Bienenstock, Shoham, & Donoghue, 2002; Shenoy et al., 2003; Taylor, Tillery, & Schwartz, 2002; Townsend, Subasi, & Scherberger, 2011; Velliste, Perel, Spaulding, Whitford, & Schwartz, 2008). Intrinsic decoding, without need of labeling, promises powerful novel approaches to BMIs that should be insensitive to instability or specific activity changes in individual neurons. In sum, intrinsic methods should prove consequential for issues of neural population representation and decoding in the various fields of cognitive neuroscience.

Appendix A: Historical Roots of Population Coding

Population coding originated in the eighteenth century with the development of trichromatic theories of color vision (Mollon, 2003; Weale, 1957). Lomonosov (1756) and Palmer (1777) both proposed that responses of just three classes of retinal receptors were sufficient to produce the percept of all colors, an idea that became more widely disseminated following its presentation by Young (1802). Helmholtz (1909/1962), in his *Treatise on Physiological Optics*, originally published in 1860, elaborated on Young's proposal, providing a schematic set of tuning curves for population coding of color (see Figure 7).

Helmholtz also contributed the first model for decoding a population in his line element theory. Created to explain color discrimination data, Helmholtz's theory treated the representation of each perceived color as a point in a three-dimensional space, given by activations of the three color channels. Under the theory, two colors became discriminable when the Euclidean distance between their 3D representations reached a certain threshold. Over the years, more elaborate versions of this model have been

developed to account for a growing body of psychophysical data (Vos, 1979; Wyszecki & Stiles, 1982). Helmholtz's line element model decodes populations based on the response difference between two stimuli. It is therefore an example of dealing with populations in an unlabeled intrinsic manner to produce relational coding rather than in an extrinsic manner to produce atomistic coding of individual stimuli.

Within visual psychophysics, starting in the late 1960s, there was an upswing of interest in population coding, expanding from the original color models to a variety of other visual variables (Campbell & Robson, 1968; Levinson & Sekuler, 1975; Sachs, Nachmias, & Robson, 1971; Wilson & Bergen, 1979; Wilson & Gelb, 1984). Population coding was reviewed from the perspective of psychophysical theory by Thomas (1985). In addition to the line element model, he presented two other models adopting an extrinsic approach, weighted averaging of tuning curve peaks and maximum likelihood estimation (MLE) (see appendix B). The later two methods would be independently developed within neurophysiology. Thomas emphasized the labeled nature of peak averaging and MLE; this is not always made explicit within the neurophysiology literature. (See Rose, 1999, for a broader perspective on labeling in psychophysical theories.)

Independent of work in psychophysics, population coding ideas were also developed within computer science under the name of parallel distributed processing (PDP), involving connectionist modeling of networks with neural-like elements (Feldman & Ballard, 1982; Hinton, 1981; Hinton, McClelland, & Rumelhart, 1986). Early PDP models that used neural network learning algorithms to create population codes involved studies of motion processing to solve the aperture problem (Sereno, 1987, 1993; see Figure 8), shape from shading (Lehky & Sejnowski, 1988, 1990), eye position gain fields (Zipser & Andersen, 1988), and the vestibular-ocular reflex (Anastasio & Robinson, 1989).

However, the PDP research program was not heavily concerned with developing explicit decoding algorithms for populations with some exceptions (e.g., Sereno, 1987, 1993). More typically within PDP modeling, population responses were fed into other populations without there ever being a need to explicitly assign interpretations to patterns of activity within intermediate layers of a network (Feldman & Ballard, 1982). Although the input and output layers of supervised learning PDP models and the input layers of unsupervised learning PDP models (e.g., Sereno & Sereno, 1991) are labeled and hence extrinsic, it is important to note that labeling model neurons per se for training purposes does not exclude using intrinsic methods to interpret the intermediate or output layer population activity of trained networks.

Although not primarily oriented toward decoding methods, the PDP work did raise the profile of population coding ideas within neurophysiology during the 1980s. This neurophysiological work was very concerned with interpreting population activities found in experimental data. A good

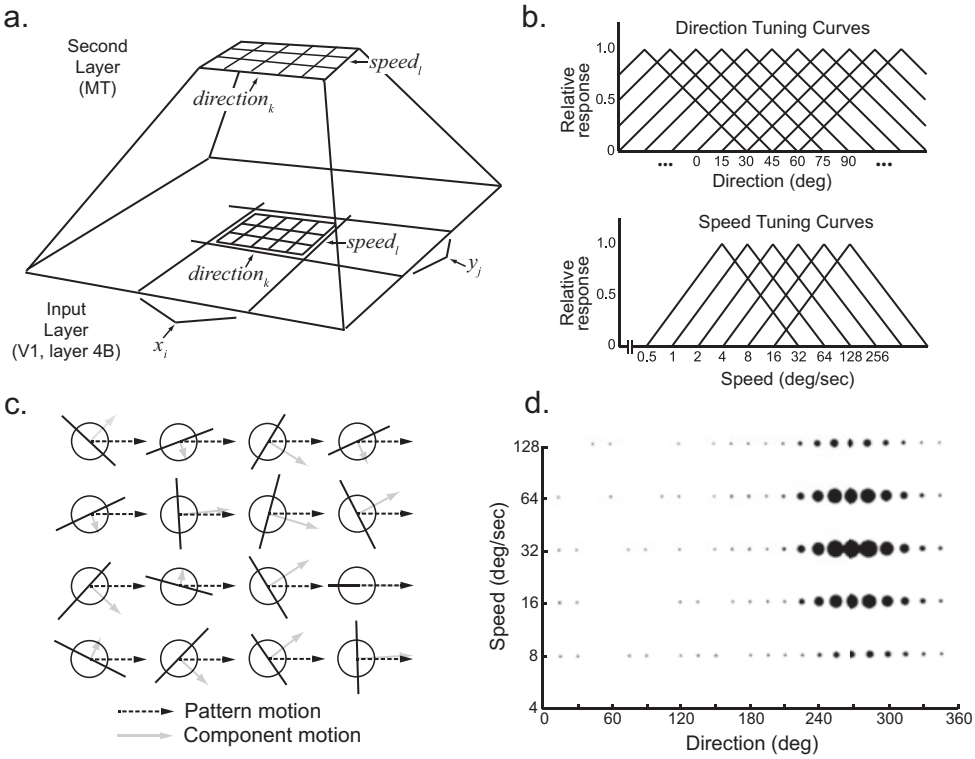


Figure 8: Early connectionist model of visual motion (Sereno, 1987, 1993) that solves the aperture problem (see Movshon, Adelson, Gizzi, & Newsome, 1985), using extrinsic (labeled) population coding. (a) Input layer (V1) neurons sensitive to local component of motion. Neurons have 2D tuning curves sensitive to speed and direction, with the component motion population replicated at different spatial locations. Output-layer (MT) neurons indicate a global pattern of motion. (b) 1D cross-sections of 2D tuning curves. (c) Example pattern for training network. The desired output is a rightward pattern motion (dashed black arrow). Different input neurons are sensitive to motion component (gray arrow) perpendicular to local feature within circular RF (aperture). (d) Activation of output units for pattern moving with direction 270 degrees and speed 32 deg/sec, after network training. Positions of black dots correspond to tuning curve peaks; diameters indicate activation levels. (Adapted from illustrations in Anderson, 1995; Sereno, 1993.)

demonstration of population coding in neurophysiology using extrinsic methods involves place cells in the hippocampus (Wilson & McNaughton, 1993). Each place cell is active when the animal is located in a particular region of the physical environment, and the joint activity of a population of such cells gives a fairly precise determination of the animal's location. Other examples of population encoding include the generation of the direction and magnitude of saccadic eye movements by the superior colliculus

(Lee, Rohrer, & Sparks, 1988), the control of the direction of arm movements in motor cortex (Georgopoulos, Schwartz, & Kettner, 1986), and the encoding of visual motion in cortical area MT (Groh, Born, & Newsome, 1997), all using extrinsic approaches.

Appendix B: Extrinsic Methods for Population Coding

B.1 Weighted Tuning Curve Averaging. This approach, also known as basis function averaging, estimates stimulus value by calculating the weighted average of population tuning curves (Pouget et al., 2000; Pouget, Dayan, & Zemel, 2003). The weights used are equal to the response (activation level) defined by the tuning curve for each neuron. For a population of tuning curves $f_i(s)$ whose values for a particular stimulus s_0 are given by $r_i = f_i(s_0)$, the weighted average curve is

$$\overline{f}(s) = \frac{\sum_{i=1}^{n} r_i f_i(s)}{\sum_{i=1}^{n} r_i}. \tag{B.1}$$

The estimated stimulus value $\hat{s}$ is then given by the value of s where the average curve $\overline{f}(s)$ has its peak.

An example of tuning curve averaging is given in Figure 9. Note that to use this technique, all neurons in the population must be labeled with parametric descriptions of their tuning curves.

B.2 Weighted Peak Averaging. Peak averaging is similar to tuning curve averaging, but instead of averaging entire tuning curves, only peak values are used. Again, averaging is weighted by the response r_i corresponding to each tuning curve. Thus, if the stimulus values corresponding to tuning curve peaks are denoted by p_i, the weighted average of the peaks is denoted by

$$\hat{s} = \frac{\sum_{i=1}^{n} r_i p_i}{\sum_{i=1}^{n} r_i}. \tag{B.2}$$

This weighted average of peaks directly gives the estimated value of the stimulus, $\hat{s}$. Figure 10a shows an example of interpreting population activity based on peak averaging. This technique assumes the values of tuning curve peaks are labeled. Georgopoulos (1995; Georgopoulos et al., 1986) was seminal in introducing peak averaging models to neurophysiology, and a number of theoretical papers and reviews cover this approach in detail (Salinas & Abbott, 1994; Sanger, 2003; Seung & Sompolinsky, 1993; Vogels, 1990; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998).

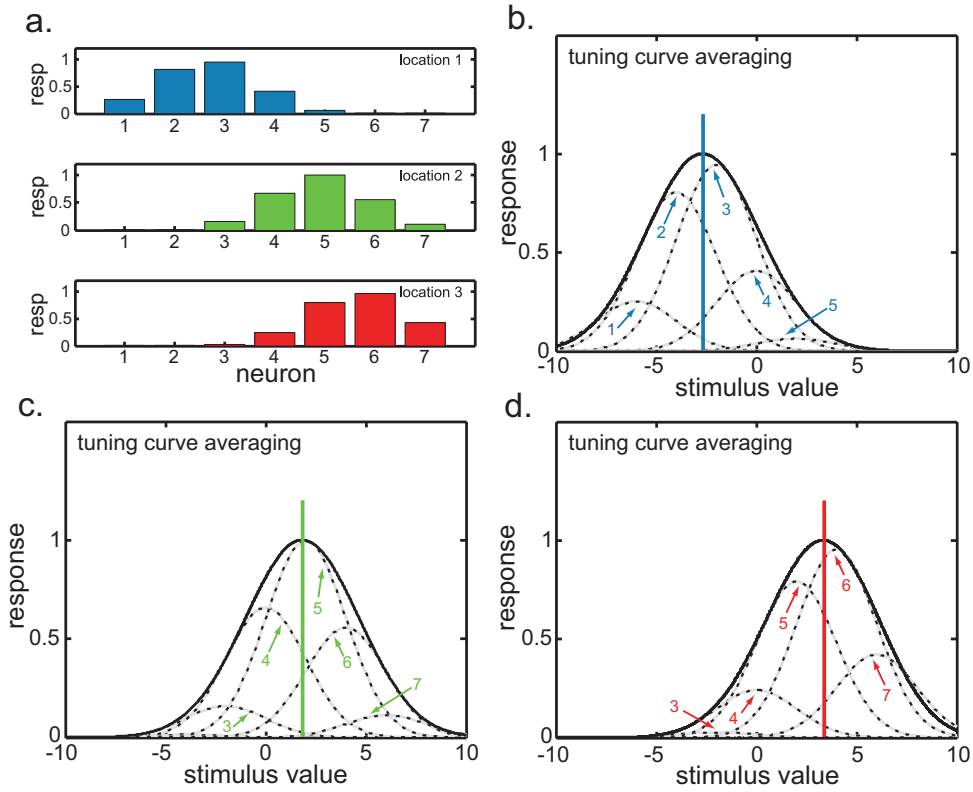


Figure 9: Weighted average of tuning curves. Example based on tuning curves and stimuli in Figure 1. (a) Responses of seven neurons to the three stimuli. (b–d) Interpreting population activity. Dashed line curves are stimulus tuning curves for individual neurons. Curve heights are proportional to the activation of each neuron as indicated in panel a. Solid curve is a weighted average of tuning curves, with height normalized to one for display purposes. The peak of the weighted average curve closely matches stimulus value (colored line).

B.3 Bayesian Estimation. The Bayesian approach codes variables in a probabilistic manner (Abbott, 1994; Brown, Frank, Tang, Quirk, & Wilson, 1998; Földiák, 1993; Oram et al., 1998; Pouget et al., 2000, 2003; Quian Quiroga & Panzeri, 2009; Sanger, 2003; Seung & Sompolinsky, 1993; Zhang et al., 1998), taking into account noise in neural tuning curves. Bayes' rule defines the following relationship among stimulus and response probabilities for the ith neuron in the population:

$$p(s|r_i) \propto p(r_i|s)p(s). \tag{B.3}$$

The output of Bayes' rule is a probability density function $p(s|r_i)$, the posterior probability. This curve indicates the probability that stimulus s has

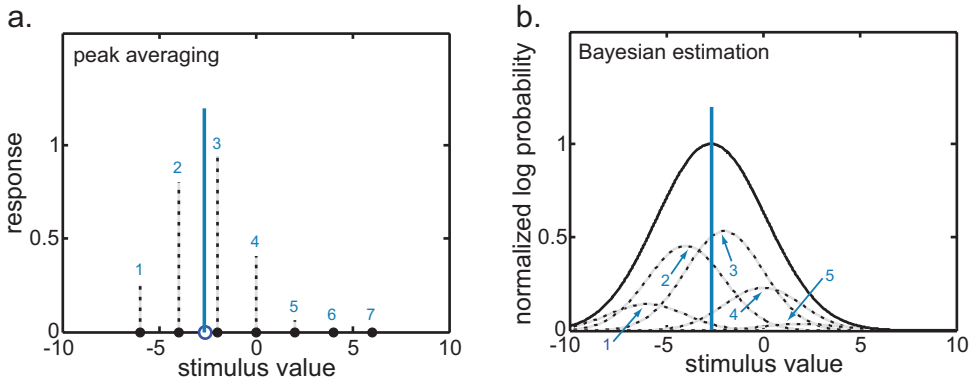


Figure 10: (a) Weighted average of tuning curve peaks. Black dots on the x-axis indicate peak values of the seven tuning curves. The stimulus response of each neuron (weighing term) is indicated by the height of the dashed line above each black dot. The weighted average of tuning curve peaks is shown by the white circle on the x-axis, indicating interpretation attached to population activity. The blue line marks physical stimulus value (the height of the blue line has no significance). (b) Bayesian estimation. Dashed lines are logs of normalized likelihood curves for individual neurons, and the solid curve is the overall population $\log(p(s|r))$ calculated by summing individual curves plus log prior probability (see equation B.5). The interpretation assigned to the population activity is stimulus value s at solid curve peak. The blue line marks actual stimulus value (the height of the blue line has no significance).

occurred as a function of neural response r_i. Applying Bayes' rule involves two probability distributions. The first is the likelihood function $p\left(r_i|s\right)$, which is the probability of r_i given s. Often when Bayesian estimation is applied, the likelihood function for each neuron is not measured but is derived from the assumption of Poisson noise statistics. The second distribution is $p(s)$, the prior probability. That denotes the probability of s occurring in the environment. Equation B.3 omits a normalization factor, which can be ignored because it affects only the height of $p(s|r_i)$, not its peak location or shape.

Once we have a probability function for each neuron, those functions are multiplied together for all neurons (assuming independent noise) to provide the joint probability across the population that s has occurred:

$$p(s|r) \propto p(s) \prod_{i=1}^{n} p(r_i|s). \tag{B.4}$$

The interpretation of population activity is then given by the stimulus value s that maximizes $p(s|r)$ (i.e., the peak of that curve). Rather than doing

this multiplication, it is convenient to take logarithms, thereby converting multiplication to addition:

$$\log[p(s|r)] \propto \log[p(s)] + \sum_{i=1}^{n} \log[p(r_i|s)]. \tag{B.5}$$

Taking logarithms changes the heights of all curves but not their peak locations. Therefore, it does not affect the stimulus value assigned to the population activity.

Figure 10b shows an example of Bayesian estimation using log-transformed probability curves. The dashed lines are $\log[p(r_i|s)]$ curves for individual neurons, and the solid curve is the overall population $\log[p(s|r)]$ calculated by summing the individual curves plus $\log[p(s)]$. The interpretation assigned to the population activity is the value of stimulus s at the peak of the summed curve.

To apply Bayesian estimation, each neuron must be labeled with descriptions of its noise properties, and the prior probabilities of stimuli in the environment must be known as well. Even if statistical estimation of stimulus values is implemented as a neural network that filters out noise (Deneve, Latham, & Pouget, 1999), that network converts a statistical estimation problem into a vector averaging or basis function problem in which neurons must still be labeled.

B.4 Maximum Likelihood Estimation. Maximum likelihood estimation is similar to Bayesian estimation, except that the prior probability $p(s)$ in equation B.3 is assumed to be constant for all stimuli s (all possible stimuli in the environment are uniformly distributed). Since $p(s)$ is constant, that leaves the likelihood function $p(r_i|s)$ as the only factor in equation B.3 that needs to be considered. Population activity is therefore interpreted as representing the stimulus value s that maximizes the likelihood function. Obviously every neuron in the population must be labeled with its likelihood function in order to use this method.

B.5 Probabilistic Population Coding. This approach attempts to represent not just the value of a variable, as was done in the methods above, but its uncertainty as well. While classic Bayesian and MLE methods treat randomness in neural responses as a nuisance variable, here it plays a fundamental role in representing stimulus uncertainty. Anderson (1994) and Földiák (1993) were early proponents of the idea that neurons may be encoding the entire probability distribution of a variable and not just its expected value. This idea has been incorporated into a variety of population coding models, including those of Ma, Beck, Latham, and Pouget (2006) and Zemel, Dayan, and Pouget (1998). Probabilistic population coding models

fall in the category of extrinsic representations because they require labeled neurons.

Appendix C: Philosophical Foundations of the Extrinsic/Intrinsic Distinction

The distinction between intrinsic and extrinsic representations in population coding finds its antecedents in different philosophical views on the nature of representation. British empiricist philosophers advocated psychological atomism, in which complex percepts were built up through associations of freestanding, independent, simple sensations. For example, Hume (1739) described the percept of a table as "impressions of colored points, disposed in a certain manner." This early commitment to atomism extended its influence into the logical atomism of Russell and Wittgenstein, which were important sources for the development of Anglo-American analytical philosophy of the twentieth century. Extrinsic population coding, with its in-built adherence to atomism, relates to this viewpoint on representations.

Following a different line of thought, development of Gestalt ideas in Germany in the late nineteenth and early twentieth centuries provided a psychological theory in which perception was irreducibly relational, not atomistic (Köhler, 1947/1992). The Gestalt perspective strongly influenced phenomenological philosophers on the European continent, most notably Merleau-Ponty. Merleau-Ponty (1964) took the opposite viewpoint from Hume's atomism, saying, "We observe at once that it is impossible, as has often been said, to decompose a perception, to make it into a collection of sensations, because in it the whole is prior to the parts." Intrinsic representation within population coding connects with this strain of thought, as sensations are coded relationally.

The distinction between atomistic and relational representations also finds antecedents in the nineteenth century with the contrasting viewpoints of the structuralist and functionalist schools of psychology (Boring, 1950).

Appendix D: Dimensionality Reduction

A neural population encodes stimuli in a high-dimensional space, where the dimensionality of the representation is equal to population size. For a population of 1000 neurons, responses to different stimuli can be represented as a set of points in a 1000-dimensional space. Typically the points would not be expected to form a uniform cloud within the high-dimensional space but show internal structure. That structure will depend on the relationships between stimuli—for example, how different stimuli are similar in some respects but dissimilar in others.

Figure 11 gives an example of internal structure within a set of points in a very small population ($n = 3$) for purpose of illustration. Each point represents a different stimulus encoded by a three-neuron population. In

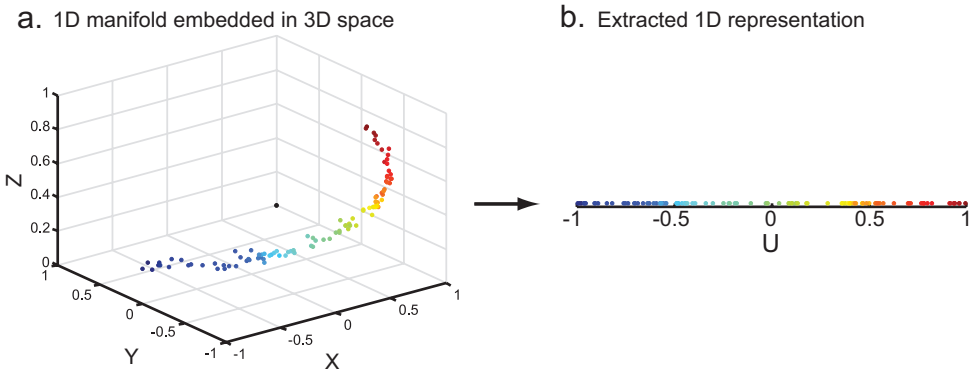


Figure 11: Example of dimensionality reduction. Intrinsic methods for interpreting population activity typically include a dimensionality reduction procedure. (a) One-dimensional manifold embedded in a three-dimensional encoding space (neural population of three neurons). Each dot represents a population response to a different stimulus. (b) One-dimensional representation space after dimensionality reduction. Colors have no significance other than to assist in comparing patterns across panels.

Figure 11a, although all the points are in 3D space, they are largely restricted to lying along a smooth 1D subspace, or manifold. With a dimensionality-reduction procedure, representation of the points can be transformed from 3D to 1D (see Figure 11b), while conserving essential relationships among the points. Possible benefits of low-dimensional manifolds embedded in high-dimensional neural representations for perceptual processing are discussed by Lehky and Sereno (2011), Sereno and Lehky (2011), and Seung and Lee (2000).

In this simple 3D example, the low-dimensional structure in the stimulus representations is obvious even without performing dimensionality reduction. However, in a more realistic situation where the stimuli were embedded in a high-dimensional space, the existence of low-dimensional manifolds would not be apparent without performing a dimensionality-reduction procedure.

In addition to the multidimensional scaling discussed in the main text, additional dimensionality-reduction algorithms that have been used to create intrinsic representations within neuroscience include principal component analysis (Lehky & Sereno, 2011) and multiple discriminant analysis (Lin, Osan, & Tsien, 2006). There are many other dimensionality-reduction algorithms, each with different characteristics (Lee & Verleysen, 2007; van der Maaten, Postma, & van den Herik, 2009). Applications of these new techniques to intrinsic coding in neural populations remain to be explored.

Acknowledgments

We thank Wei Ji Ma, Saumil Patel, Stuart Red, and Anthony Wright for comments on the manuscript. This research was supported in part by NSF grant 092436 to A.B.S. and NIDA grant R21DA024293 to M.E.S.

References

Abbott, L. F. (1994). Decoding neuronal firing and modelling neural networks. *Quarterly Reviews of Biophysics*, *27*, 291–331.

Anastasio, T. J., & Robinson, D. A. (1989). The distributed representation of vestibulo-oculomotor signals by brain-stem neurons. *Biological Cybernetics*, *61*, 79–88.

Anderson, C. H. (1994). Basic elements of biological computational systems. *International Journal of Modern Physics C*, *5*, 135–137.

Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*, 358–366.

Bokil, H. S., Pesaran, B., Andersen, R. A., & Mitra, P. P. (2006). A method for detection and classification of events in neural activity. *IEEE Transactions on Biomedical Engineering*, *53*, 1678–1687.

Borg, I., & Groenen, P. (2010). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York: Springer.

Borghuis, B. G., Ratliff, C. P., Smith, R. G., Sterling, P., & Balasubramanian, V. (2008). Design of a neuronal array. *Journal of Neuroscience*, *28*, 3178–3189.

Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). New York: Prentice Hall.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., & Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, *18*, 7411–7425.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, *197*, 551–566.

Churchland, P. M. (2012). *Plato's camera: How the physical brain captures a landscape of abstract universals*. Cambridge, MA: MIT Press.

Churchland, P. S., & Churchland, P. M. (2002). Neural worlds and real worlds. *Nature Reviews Neuroscience*, *3*, 903–907.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, *2*, 740–745.

DeVries, S. H., & Baylor, D. A. (1997). Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *Journal of Neurophysiology*, *78*, 2048–2060.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, *21*, 449–498.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Edelman, S., & Intrator, N. (1997). Learning as extraction of low-dimensional representations. In D. Medin, R. Goldstone & P. Schyns (Eds.), *Mechanisms of perceptual learning*. New York: Academic Press.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, *6*, 205–254.

Földiák, P. (1993). The "ideal homonculus": Statistical inference from neural population responses. In F. H. Eekman& J. M. Bower (Eds.), *Computation and neural systems* (pp. 55–60). Norwell, MA: Kluwer.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*, 1187–1196.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*, 343–346.

Gao, Y., Black, M. J., Bienenstock, E., Shoham, S., & Donoghue, J. P. (2002). Probabilistic inference of hand motion from neural activity in motor cortex. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems*, *14* (pp. 221–228). Cambridge, MA: MIT Press.

Georgopoulos, A. P. (1995). Motor cortex and cognitive processing. In M. S. Gazzaniga (Ed.), *The cognitive sciences* (pp. 505–517). Cambridge, MA: MIT Press.

Georgopoulos, A. P., Caminiti, R., Kalaska, J. F., & Massey, J. T. (1983). Spatial coding of movement: A hypothesis concerning the coding of movement direction by motor cortical populations. *Experimental Brain Research*, *7* (Suppl.), 327–336.

Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, *2*, 1527–1537.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, *233*, 1416–1419.

Gochin, P. M. (1996). The representation of shape in the temporal lobe. *Behavioural Brain Research*, *76*, 99–116.

Groh, J. M., Born, R. T., & Newsome, W. T. (1997). How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *Journal of Neuroscience*, *17*, 4312–4330.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Helmholtz, H.L.F. v. (1962). *Treatise on physiological optics* (3rd ed.) (J.P.C. Southall, Trans.). New York: Dover. (Original work published 1909.)

Hinton, G. E. (1981). *Shape representation in parallel systems.* Paper presented at the Seventh International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*, 428–434.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 77–109). Cambridge, MA: MIT Press.

Hume, D. (1739). *A treatise of human nature, Book I, Part II, Section III*. Kindle ed.

Jeannerod, M., & Jacob, P. (2005). Visual cognition: A new look at the two-visual systems model. *Neuropsychologia*, *43*, 301–312.

Kayaert, G., Biederman, I., & Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cerebral Cortex*, *15*, 1308–1321.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*, 4296–4309.

Köhler, W. (1992). *Gestalt psychology: An introduction to new concepts in modern psychology*. New York: Liveright. (Original work published 1947.)

Kourtzi, Z., & Connor, C. E. (2011). Neural representations for object perception: Structure, category, and adaptive coding. *Annual Review of Neuroscience*, *34*, 45–67.

Krakauer, J. W., Pine, Z. M., Ghilardi, M.-F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, *20*, 8916–8924.

Kriegeskorte, N. (2009). Relating population-code representations between man, monkey, and computational models. *Frontiers in Neuroscience*, *3*, 363–373.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141.

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, *332*, 357–360.

Lee, J. A., & Verleysen, M. (2007). *Nonlinear dimensionality reduction*. New York: Springer.

Lehky, S. R., & Sejnowski, T. J. (1988). Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature*, *333*, 452–454.

Lehky, S. R., & Sejnowski, T. J. (1990). Neural network model of visual cortex for determining surface curvature from images of shaded surfaces. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *240*, 251–278.

Lehky, S. R., & Sereno, A. B. (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. *Journal of Neurophysiology*, *97*, 307–319.

Lehky, S. R., & Sereno, A. B. (2011). Population coding of visual space: Modeling. *Frontiers in Computational Neuroscience*, *4*, 155. doi:110.3389/fncom.2010.00155

Levinson, E., & Sekuler, R. (1975). The independence of channels in human vision selective for direction of movement. *Journal of Physiology*, *250*, 347–366.

Lin, L., Osan, R., & Tsien, J. Z. (2006). Organizing principles of real-time memory encoding: Neural clique assemblies and universal neural codes. *Trends in Neurosciences*, *29*, 48–57.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, *5*, 552–563.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Lomonosov, M. H. (1756). *Oratio de origine lucis novam theoriam colorum*. Petropoli: Typis Academie Scientiarum.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

Merleau-Ponty, M. (1964). *The primacy of perception* (J. M. Edie, Trans.). Evanston, IL: Northwestern University Press.

Mollon, J. (2003). The origins of modern color science. In S. Shevell (Ed.), *Color science* (pp. 1–39). Washington, DC: Optical Society of America.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass & C. G. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117–151). Vatican City: Pontificia Academiae Scientiarum.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology*, *83*, 2580–2601.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, *4*, 1244–1252.

Oram, M. W., Földiák, P., Perrett, D. I., & Sengpiel, F. (1998). The "Ideal Homunculus": Decoding neural population signals. *Trends in Neurosciences*, *21*, 259–265.

Palmer, G. (1777). *Theory of colours and vision*. London: Leacroft.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*, 125–132.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*, 381–410.

Pouget, A., Deneve, S., Ducom, J.-C., & Latham, P. E. (1999). Narrow versus wide tuning curves: What's best for a population code? *Neural Computation*, *11*, 85–90.

Quian Quiroga, R., & Panzeri, S. (2009). Extracting information from neuronal populations: Information theory and decoding approaches. *Nature Reviews Neuroscience*, *10*, 173–185.

Richmond, B. J., Optican, L. M., Podell, M., & Spitzer, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *Journal of Neurophysiology*, *57*, 132–146.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Rolls, E. T., & Tovée, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*, 713–726.

Rose, D. (1999). The historical roots of the theories of local signs and labelled lines. *Perception*, *28*, 675–685.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America*, *61*, 1176–1186.

Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, *1*, 89–107.

Salinas, E., & Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *Journal of Neuroscience*, *15*, 6461–6474.

Sanger, T. D. (2003). Neural population codes. *Current Opinion in Neurobiology*, *13*, 238–249.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*, 417–457.

Sereno, A. B., & Lehky, S. R. (2011). Population coding of visual space: Comparison of spatial representations in dorsal and ventral pathways. *Frontiers in Computational Neuroscience*, *4*, 159. doi:110.3389/fncom.2010.00159

Sereno, M. E. (1987). *Implementing stages of motion analysis in neural networks*. Paper presented at the Proceedings of the Ninth Annual Conference of the Cognitive Science Society.

Sereno, M. E. (1993). *Neural computation of pattern motion: Modeling stages of motion analysis in the primate visual cortex*. Cambridge, MA: MIT Press.

Sereno, M. I., & Sereno, M. E. (1991). Learning to see rotation and dilation with a Hebb rule. In R. P. Lippmann, J. Moody & D. S. Touretzky (Eds.), *Advances in neural information processing systems, 3* (pp. 320–326). San Mateo, CA: Morgan Kaufmann.

Seung, H. S., & Lee, D. D. (2000). Cognition. The manifold ways of perception. *Science*, *290*, 2268–2269.

Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 10749–10753.

Shenoy, K. V., Meeker, D., Cao, S., Kureshi, S. A., Pesaran, B., Buneo, C. A., et al. (2003). Neural prosthetic control signals from plan activity. *Neuroreport*, *14*, 591–596.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390–398.

Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, *1*, 1–17.

Shepard, R. N., & Podgorny, P. (1978). Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 189–237). Hillsdale, NJ: Erlbaum.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, *66*, 170–189.

Taylor, D. M., Tillery, S.I.H., & Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science*, *296*, 1829–1832.

Thomas, J. P. (1985). Detection and identification: how are they related? *Journal of the Optical Society of America A. Optics and Image Science*, *2*, 1457–1467.

Townsend, B. R., Subasi, E., & Scherberger, H. (2011). Grasp movement decoding from premotor and parietal cortex. *Journal of Neuroscience*, *31*, 14386–14398.

van der Maaten, L., Postma, E., & van den Herik, J. (2009). *Dimensionality reduction: A comparative review* (Tech. Rep. TiCC-TR 2009-005). Tilburg, Netherlands: Tilburg University Centre for Creative Computing.

Velliste, M., Perel, S., Spaulding, M. C., Whitford, A. S., & Schwartz, A. B. (2008). Cortical control of a prosthetic arm for self-feeding. *Nature*, *453*, 1098–1101.

Vogels, R. (1990). Population coding of stimulus orientation by striate cortical cells. *Biological Cybernetics*, *64*, 25–31.

Vos, J. J. (1979). Line elements and physiological models of color vision. *Color Research and Application*, *4*, 208–216.

Wallman, J., & Fuchs, A. F. (1998). Saccadic gain modification: Visual error drives motor adaptation. *Journal of Neurophysiology*, *80*, 2405–2416.

Weale, R. (1957). Trichromatic ideas in the seventeenth and eighteenth centuries. *Nature*, *179*, 648–651.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, *19*, 19–32.

Wilson, H. R., & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America A. Optics and Image Science*, *1*, 124–131.

Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, *261*, 1055–1058.

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae* (2nd ed.). New York: Wiley.

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.*, *11*, 1352–1360.

Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, *256*, 1327–1331.

Young, T. (1802). The Bakerian lecture: On the theory of light and colours. *Philosophical Transactions of the Royal Society of London*, *92*, 12–48.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, *10*, 403–430.

Zhang, K., Ginzburg, I., McNaughton, B. L., & Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology*, *79*, 1017–1044.

Zhang, K., & Sejnowski, T. J. (1999). Neural tuning: To broaden or to sharpen. *Neural Computation*, *11*, 75–84.

Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, *331*, 679–684.

Received September 12, 2012; accepted March 12, 2013.