12

Coding in Large-Scale Cortical Populations

Jason N. MacLean and Nicholas G. Hatsopoulos

Abstract

Theories of information coding in cortical populations have been put forth for many years, but only recently have experimental methods become available to permit simultaneous recordings from hundreds of neurons, thus allowing these theories to be tested. This chapter discusses some of the more prominent theories and argues that they fall along a spectrum of coding schemes, ranging from population codes that are built up from single-neuron tuning functions to codes that emerge from the collective dynamics of cortical populations. At the extremes, these theories are incompatible: one relies on single neurons whereas the other ingrains coarse neuronal activity into low-dimensional trajectories that summarize the covariance of activity across multiple neurons. It is proposed that both can be reconciled using a hierarchical coding scheme where relevant information is represented at the level of large-scale spatiotemporal patterns, and both individual neurons and the temporal interrelationships convey information. Antecedents to this contemporary theory can be seen in Donald Hebb's assembly phase sequences (Hebb 1949): information is encoded at the single-neuron level in terms of tuning functions, but spatiotemporal patterning of individual neurons provides context to interpret the population code fully. Moreover, the encoding perspective proposed here explicitly incorporates the synaptic implementation of the code, thus strengthening the postulate.

Introduction

Thanks to the ongoing development of technologies, such as multielectrode arrays and optical imaging, it has become increasingly routine to record simultaneously from hundreds, and soon even thousands, of neurons in awake-behaving animals. From this perspective, we begin by arguing *why* the cortex needs to use a population scheme to encode sensory, cognitive, and motor information. We then proceed to describe different population schemes and argue that it is necessary to understand *how* as well as *what* is encoded in the population to fully delineate the nature of the cortical population

code: How is the code implemented mechanistically from a cellular and synaptic perspective? What is being represented, and for which purpose, at any moment in time by the active cortical neuronal population? From an encoding perspective, the issues of *how, what,* and *for which purpose* must be established to permit a complete theory. From a decoding perspective, it is possible to black box the *how* and *for which purpose* and instead establish only *what* is being encoded.

Population coding schemes can be considered on a spectrum of neural representations where the extremes are seemingly mutually exclusive. At one end of the spectrum, population codes are built up from the sum of activity of a collection of single neurons, which we term "single neuron level" (SNL) codes. SNL codes assume that sensory, cognitive, or motor information resides at the single-neuron level in the form of tuning curves, for example. Tuning curves represent the trial-averaged response of a single neuron and reflect the fact that a neuron is more likely to spike when a specific statistical feature is present in a stimulus, when a percept, thought, or decision is generated, or preceding a specific motoric feature. The population, in turn, codes information by aggregating or combining information across neurons. A pooling code is perhaps the simplest example of an SNL code in which information is represented as a weighted average of activities of single neurons. At the other end of the spectrum, information is encoded at the neural ensemble level (NEL). One example of an NEL code assumes that single neurons in isolation are uninformative but collectively they serve as cogs in a multineuronal machine whose spatiotemporal dynamics carry information. The characterization of motor cortex as a dynamical system put forth by Shenoy and Churchland exemplifies such a code (Churchland et al. 2012; Shenoy et al. 2013). However, we subscribe to the viewpoint that the cortex employs NEL codes that do not ignore single neurons and their tuning properties while simultaneously employing higher-order statistical relationships among neurons as a coding element. Changes in both rate and spiking correlations, in response to sensory stimuli or in relation to motor output, are generally linked (de la Rocha et al. 2007; see however, Biederlack et al. 2006). To understand what we mean by this perspective, consider the analogy of our decimal system for representing numbers. Single digits, such as "1" and "2," have distinct meanings and, therefore, code information. However, "12" refers to something quite different than the two isolated digits that comprise it and yet the ensemble code depends on the meanings of the individual digits. This suggests a hierarchical coding scheme where information is carried at multiple levels of a hierarchy.

Why Use a Population Coding Scheme?

At the cortical level, it is well-established that single neurons are highly variable from moment to moment and trial to trial, even under well-controlled

conditions (Heggelund and Albus 1978; Tolhurst et al. 1983; Vogels et al. 1989; Britten et al. 1993; Arieli et al. 1996; Lin et al. 2015). Moreover, neurons are interconnected by large numbers of dynamic and generally weak synapses and consequently must integrate multiple inputs to spike. It is likely that cortical codes depend on the collective activity of large groups of neurons. Fortunately, we now have the technological ability to experimentally evaluate population coding schemes. As we discuss below, it remains unclear what the numerical size and spatial extent of an encoding population is beyond the fact that it is more than a single neuron.

SNL Codes

The population vector, one of the best known SNL codes, has proven effective in decoding reach direction from a group of motor cortical (M1) neurons (Georgopoulos et al. 1986). By fitting directional tuning curves of individual M1 neurons to cosine functions, the preferred directions of these cells are computed and represented as two- or three-dimensional vectors. The population vector is then calculated as a weighted vector sum of the preferred directions, each weighted by its firing rate. This approach can be extended to incorporate the sum of population vectors (Gilbert and Wiesel 1990; Vogels 1990). Although originally applied to sequentially recorded neurons, this approach has been extended to simultaneously recorded neurons with good decoding success (Taylor et al. 2002). Population vector decoding has been applied to a variety of systems, including the visual system, to estimate the orientation of faces (Oram et al. 1998) and, in the auditory system, to localize a sound source based on intra-aural time differences (Fitzpatrick et al. 1997).

In general, the population vector decoder does not provide an optimal estimator in the sense that it does not always minimize the variance of the estimator (Deneve et al. 1999); this can result in poor decoding performance (Montijn et al. 2016). Other SNL population coding schemes have been shown to be optimal under certain assumptions, including the optimal linear estimator (OLE), the indirect OLE, and maximum likelihood estimation (Salinas and Abbott 1994; Deneve et al. 1999; Wang et al. 2007b). The indirect OLE approach has been extended for real-time, closed-loop neural prosthesis control of a 10-dimensional robot in a paralyzed patient (Wodlinger et al. 2015). In this application, single-neuron tuning models included not only reach direction but also wrist orientation and grasp velocities. Maximum likelihood estimation can also be implemented using a recurrent network model under certain assumptions (Deneve et al. 1999).

Despite the general success of the population vector and other near-optimal estimators, an important assumption underlies these population decoding schemes: the firing rates of the population are assumed to be statistically

independent, conditioned on the stimulus, decision, or movement. It is this assumption, in fact, which provides the population vector scheme with its ability to reduce noise by averaging over neurons, which in turn makes it particularly effective for decoding (Zohary et al. 1994). However, trial-to-trial correlations in spike counts that are independent of the specific stimulus or movement (i.e., noise correlations) have been documented in many sensory and motor cortical areas and are often considered to be a nuisance for SNL population codes because the variance of a population mean estimator will not decrease as 1/N, where N is the number of neurons in the population. More generally, however, noise correlations do not always reduce population-level information and, in fact, can improve stimulus discrimination in population activity (Poort and Roelfsema 2009). Theoretical work by the Pouget group has identified a specific kind of noise correlation (differential correlations) that is proportional to the product of the derivative of the tuning curves and impacts the information capacity of SNL population codes (Moreno-Bote et al. 2014). The assumption of independence makes it difficult to extend this framework to an encoding model that incorporates synaptic and cellular mechanisms, since neurons are interconnected and collectively drive and support activity in the active population; that is, they are not independent (Renart et al. 2010). However, the utility of SNL codes makes clear that single neurons do indeed represent information. We suggest that this fact should be incorporated into any cortical population coding scheme.

NEL Codes

Regardless of the cortical area recorded or the recording method, researchers have found evidence of structured spatiotemporal activity in all cases: this is consistent with the Hebbian assembly phase sequence, which postulates that across trials, functionally related assemblies or ensembles of neurons reliably propagate spikes from one ensemble to another (Hebb 1949; Abeles and Gerstein 1988; Villa et al. 1999; Beggs and Plenz 2004; Gourevitch and Eggermont 2010; Bathellier et al. 2012; Gansel and Singer 2012; Palm et al. 2014; Peters et al. 2014; Reyes-Puerta et al. 2015). This coding scheme is made more attractive by the fact that it easily allows for inclusion of cell identity and is consequently compatible with cellular and synaptic mechanisms (Kruskal et al. 2013).

Correlation Codes

While it is widely recognized that cortical neurons exhibit correlated firing at different timescales and constrain population activity patterns, we postulate that these correlations code information.

Synchrony

Synchronous firing of neuron pairs at the millisecond timescale have been shown to signal a variety of different sensory and motor states. Pairs of oscillating neurons in visual cortex have been shown to synchronize when stimulus features form a single, coherent visual stimulus (Gray et al. 1989). This and other experiments have been used to support the view that synchronization serves to bind stimulus features that are perceptually related (Singer 2018a). In the motor domain, synchrony has been shown to encode movement direction and expected visual cues for initiating movement in primary motor cortex (Riehle et al. 1997; Hatsopoulos et al. 1998; Kilavik et al. 2009; Denker et al. 2011). Moreover, larger groups of synchronous M1 neurons have been shown to signal task epochs and behavioral conditions in an instructed delay reach-to-grasp task (Torre et al. 2016). By identifying so-called cortico-motoneuronal cells that make direct synaptic connections with motor neurons in the spinal cord via spike-triggered electromyography recordings, it has been found that cortico-motoneuronal pairs which share similar muscle fields tend to be more synchronized than pairs that have different projection fields (Jackson et al. 2003). Moreover, modeling work suggests that inputs to motor neurons can generate substantially more force output when they are synchronized (Baker et al. 1999). Therefore, cortical synchrony may provide a mechanism for directly affecting behavioral output.

Noise Correlations

Despite their potential detrimental effect on SNL population codes, noise correlations have been shown to signal different behavioral states. For example, spike count correlations in frontal eye fields have been shown to vary dynamically in distinct ways for different saccadic eye movements, despite the fact the neurons' firing rates do not modulate (Vaadia et al. 1995). Spike count correlations between pairs of M1 neurons have been shown to increase when two-element movement sequences were preplanned, compared to when they were planned one at a time, even when the firing rate modulations of the neurons did not differ under the two conditions (Hatsopoulos et al. 2003). This suggests that task complexity impacts the correlational structure.

Latent Variable Codes

A latent variable code is a form of NEL code that assumes there are latent variables that are not explicitly observed but inferred from spiking data (i.e., the manifest variables). Principal components analysis and factor analysis are examples of continuous latent variable methods that have been used extensively to characterize large-scale cortical recordings (Mazor and Laurent 2005; Yu et al. 2009a). Unlike pooling codes, these methods depend on the covariance

of neuronal activity to find subspaces in which most of the variance of activity resides. It is postulated that population activity resides in a much smaller subspace than what would be theoretically possible because of (a) the interconnections between neurons, (b) the minimal coding requirements for the task at hand (see below), or (c) due to common input. Regardless, these subspaces have been demonstrated to enable decoding of cortical population activity and have also provided insights into how cortex represents information. For example, animals have difficulty learning to control brain–machine interfaces when this requires generating population activity outside of these intrinsic subspaces (Sadtler et al. 2014). Moreover, a recent study has shown that choice behavior in a visual discrimination task depends on population activity that resides in a principal component subspace, even though it is suboptimal (Ni et al. 2018).

Besides serving to reduce the dimensionality of high-dimensional neural data, continuous latent variables reveal a dynamic structure in the active population that is not evident at the single-neuron level. In an elegant set of studies, Shenoy and colleagues have characterized motor and premotor ensembles as neural trajectories that exhibit relatively simple rotational dynamics during movement execution in a reduced subspace (Churchland et al. 2012; Shenoy et al. 2013). Different movements (e.g., movements to different goals) correspond to different initial conditions within the movement subspace, but all movements share these underlying rotation dynamics. Moreover, movement planning or preparation resides in an orthogonal subspace to the space that corresponds to the muscle activation and movement (Kaufman et al. 2014). According to Kaufmann et al., single neurons serve only as the building blocks of a multineuronal dynamical system that traverses different portions of state space. If anything is coded, it is the location or dynamics of the population state that represents planning or movement of a particular type. From an encoding perspective, the informative nature of structured population dynamics indicates that the spatiotemporal interactions between neurons, and the corresponding cellular and synaptic mechanisms that generate this structure, will be fundamental to any unifying theory of cortical population coding.

Hidden Markov models (HMMs) are examples of discrete latent codes that have been used to characterize cortical population dynamics as a sequence of discrete "hidden" states which account for shared variability across the population (Radons et al. 1994; Abeles et al. 1995; Seidemann et al. 1996; Kemere et al. 2008). The probability of a state transition is based solely on the current state of the system. Moreover, each hidden state has an associated probability distribution for observed spiking responses across the population. As compared to populations of poststimulus time histograms, coupling hidden states can more accurately predict the spiking statistics of population activity, particularly when activity is not time-locked to an external stimulus (Abeles et al. 1995). Moreover, each state is associated with distinct pairwise neural correlations, often reflecting different behavioral states. Recently we applied a HMM to simultaneously recorded spiking data from primary motor cortex while

monkeys engaged in reaching movements and found that population activity appears to transition among a small set of states (Kadmon Harpaz et al. 2018). More importantly, we discovered that state transitions correspond to velocity extrema of reaching, such that a given state corresponds to either an accelerative or decelerative phase of reaching in a particular direction. Simulations using single-neuron tuning models of direction and speed applied to the reaching movements could not replicate the findings of our HMM. This implies that motor cortical population dynamics may be more accurately characterized as transitions among discrete hidden states that code for accelerative/decelerative movements in particular directions.

Spatiotemporal Patterning

Most NEL coding schemes consider information encoded in the pattern of activity across a population but largely ignore the spatial layout of neurons within the population, despite the strong anisotropy of neurons and their connection likelihood. Studies using multielectrode arrays and voltage-sensitive dyes, however, have documented propagating wave activity consistent with spatiotemporal codes within the visual (Arieli et al. 1995; Prechtl et al. 1997; Roland et al. 2006; Xu et al. 2007), somatosensory and sensorimotor (Petersen et al. 2003; Ferezou et al. 2007), auditory (Song et al. 2006; Witte et al. 2007), and motor cortices (Rubino et al. 2006; Takahashi et al. 2015) as well as hippocampus (Lubenov and Siapas 2009; Patel et al. 2012; Zhang and Jacobs 2015). In the CA1 region of hippocampus, local field potential (LFP) waves propagate along the septotemporal axis, mediated by theta oscillations, and may serve to encode not only the present location of the animal but also the past and future locations (Lubenov and Siapas 2009). In primary and premotor motor cortices, LFP waves in beta oscillations propagate along the rostrocaudal and medial-lateral axes of the cortical surface, respectively. The propagating direction and speed do not vary with movement direction and thus do not appear to code for movement parameters. However, in an instructed delay paradigm, where a visual target of an upcoming movement is presented, visually evoked waves do encode upcoming movement direction in the amplitude and timing of these evoked potentials (Rubino et al. 2006). Moreover, the sequential firing of pairs of M1 neurons carries more directional information when the two neurons are oriented along the LFP wave-propagating axis (Takahashi et al. 2015).

We have documented another spatiotemporal pattern of activity in motor cortex that may be important for initiating movement (Best et al. 2017; Balasubramanian et al. 2019). Voluntary movement initiation involves the modulation of large populations of M1 neurons around movement onset. Despite knowledge of the temporal dynamics of cortical ensembles that lead to movement, the spatial structure of these dynamics, across the cortical sheet, have been largely ignored. We have shown (Best et al. 2017) that the timing in attenuation of the beta frequency oscillation amplitude, a neural correlate of

corticospinal excitability (Pfurtscheller and Lopes da Silva 1999), forms a spatial gradient across motor cortex prior to movement onset with a defined beta attenuation orientation from earlier to later attenuation times. We have also shown that a similar propagating pattern is evident in the modulation times of populations of M1 neurons. Interestingly, even though M1 neurons modulate activity during movement preparation well before movement onset, these modulation times do not exhibit spatial structure, suggesting that spatiotemporal structure in modulation may be necessary to trigger movement. It should be emphasized that such spatiotemporal patterns do not lessen the possibility that single neurons also carry information. Rather, we argue that cortex may be using a hierarchical coding scheme: large-scale patterning signals certain global aspects of behavior while single neurons code for more specific aspects or, alternatively, large-scale patterning provides context for single-neuron codes which together form a population code.

How Would an Ensemble Code Work?

To truly understand the code, we must eventually synthesize it with the cellular and synaptic mechanisms that implement the code. Neocortical microcircuitry includes the neurons and synaptic connections within volumes of ~500 μm × 500 μm × 1000 μm. Individual neurons are highly interconnected and connection likelihood is biased toward neighboring neurons (Song et al. 2005; Ko et al. 2011; Perin et al. 2011). Moreover, in primary sensory cortices, only ~5% of connections that a neuron receives arise from ascending inputs (Peters and Payne 1993; Douglas and Martin 2004), and a comparable portion originates from distal cortical regions (Budd 1998). The microcircuit, then, represents the scale over which (a) most excitatory and inhibitory interactions take place, and (b) synaptic connections strengthen or weaken according to the relative spike timing between pre- and postsynaptic neurons (Kruskal et al. 2013). This is also the scale at which Hebbian learning occurs. Consequently, microcircuits form the building blocks from which the cortical population code is built. Imaging approaches permit researchers to densely sample neurons allowing microcircuit dynamics to be more directly linked to synaptic mechanisms (Ko et al. 2011; Chambers and MacLean 2015).

Individual synaptic connections are weak, ranging from 0.2–1.0 mV (Holmgren et al. 2003), and patterns of spiking are complex and variable, which make the mapping between structure and functional dynamics far from straightforward. However, a small set of correlations in population activity is indicative of synapses that are actively involved in the recruitment of postsynaptic neurons; namely, those synaptic inputs that occur at just the right time to drive the postsynaptic neuron to threshold (Ko et al. 2011; Chambers and MacLean 2015, 2016). We found that these "recruitment" synapses are directly linked to specific higher-order motifs in population correlational structure. To

make this observation, we accounted for the nonlinear integrative properties of neurons by using a combination of spiking neuronal network models and experimental measurements (Chambers and MacLean 2016). Fan-in triangles, where two input neurons are themselves connected, coordinate the timing of presynaptic inputs during ongoing activity to facilitate postsynaptic spiking. Interplay between higher-order synaptic connectivity and the integrative properties of neurons constrains the structure of network dynamics and shapes the routing of information in neocortex (Chambers and MacLean 2016).

We have also linked higher-order correlational structure to the activity of neurons in mouse visual cortex by imaging microcircuit responses to visual input in awake, head-fixed, ambulating mice. We found that trial-averaged tuning properties of neurons explain only a small fraction of the single-trial activity of neurons, similar to other studies (Reimer et al. 2014; Montijn et al. 2016). By summarizing the dynamics as a functional network, we are able to use the neighbors of a neuron, necessarily including both tuned and untuned neurons, to predict individual neuron activity on a single-trial basis. Perception and behavior take place in real time, after all, so it is necessary that any population encoding model of stimulus representations in cortex encompass single-trial responses. Moreover, again a specific triplet motif maximized predictions of single-trial responses (Dechery and MacLean 2017). Consistent with these results from visual cortex, Meshulam et al. (2017) found that they were able to best predict single-neuron activity in hippocampus by employing a maximum entropy model that incorporated the state of all of the recorded neurons, including neurons that have well-defined place fields as well as those which do not. Together, these studies indicate that the collective behavior of neurons, both tuned and untuned, are necessary to predict single-trial neuronal responses and argue strongly in favor of an NEL coding framework.

At any given moment in time, it is unclear how many neurons are necessary to encode a stimulus or motor output. Using connectivity estimates, synaptic strength estimates, as well as membrane potentials and conductance states of individual neurons, a lower bound estimate of approximately 100 presynaptic neurons has been postulated (Ainsworth et al. 2012). As described above, low-dimensional summaries of population dynamics are a very effective means to decode population activity (Briggman et al. 2005; Churchland et al. 2007; Harvey et al. 2009). Consistently, a number of studies have found that the number of observed spatiotemporal neuronal activity patterns are limited (MacLean et al. 2005; Luczak et al. 2007, 2015; Luczak and Maclean 2012). Moreover it appears that baseline connectivity also constrains feasible dynamics during learning (Shenoy et al. 2013; Sadtler et al. 2014; Shenoy and Carmena 2014). All of these data suggest that population dynamics occupy a much lower-dimensional space than the number of neurons recorded, which indicates diminishing returns as the number of recorded neurons increases into the thousands. Two recent studies, however, provide compelling arguments that the number of neurons, and more loosely the number of dimensions, necessary

to encode a motor output or sensory stimulus likely depends on what is being encoded. The neural task complexity theory (NTC) posits that the complexity of the task that an animal is performing combined with the smoothness of the recorded neural trajectories determine the size of the population code (Gao et al. 2017). Further, NTC theory maps the behavioral parameters within neural dynamics and predicts that this volume will be small when tasks are simple and trajectories are smooth (i.e., the covariance of neuronal activity is well captured in the recorded population). NTC theory predicts that the size of the coding pool and, relatedly, the number of dimensions necessary to summarize population dynamics scale with the nature or difficulty of the task or stimulus.