**The population doctrine in cognitive neuroscience**

R. Becket Ebitz\(^1,\)\(^*\) and Benjamin Y. Hayden\(^2\)

\(^1\)Department of Neurosciences, Faculté de médecine, Université de Montréal, Montréal, QC, Canada
\(^2\)Department of Neuroscience, Center for Magnetic Resonance Research, and Center for Neuroengineering, University of Minnesota, Minneapolis, MN, USA

\(*\)Correspondence: becket@ebitzlab.com

https://doi.org/10.1016/j.neuron.2021.07.011

A major shift is happening within neurophysiology: a population doctrine is drawing level with the single-neuron doctrine that has long dominated the field. Population-level ideas have so far had their greatest impact in motor neuroscience, but they hold great promise for resolving open questions in cognition as well. Here, we codify the population doctrine and survey recent work that leverages this view to specifically probe cognition. Our discussion is organized around five core concepts that provide a foundation for population-level thinking: (1) state spaces, (2) manifolds, (3) coding dimensions, (4) subspaces, and (5) dynamics. The work we review illustrates the progress and promise that population-level thinking holds for cognitive neuroscience—for delivering new insight into attention, working memory, decision-making, executive function, learning, and reward processing.

**INTRODUCTION**

Cognition gives us the flexibility to selectively attend to a stimulus, hold information in mind, pursue arbitrary goals, implement executive control, or decide between actions by learning about and weighing beliefs about their outcomes. Neurophysiology has revealed much about the neural basis of cognition over the last fifty years, generally by examining responses in one neuron at a time. However, brain areas contain hundreds of millions of neurons (Herculano-Houzel, 2009), and somewhere between the scale of single neurons and the gross pooling in electroencephalogram (EEG) or fMRI, neuronal populations produce macroscale phenomena that are poised to link the scale of single neurons to the scale of behavior.

The term population doctrine describes the belief that the population, not the neuron, is the fundamental unit of computation (Saxena and Cunningham, 2019). This idea is not new: some researchers have always studied how neurons behave collectively, whether that was Hebb’s cell assemblies in the 1940s (Hebb, 1949) or Georgopoulos’s population vectors in the 1980s (Georgopoulos et al., 1986). Population ideas have always been influential in theoretical neuroscience, but they have sometimes lain dormant among experimentalists. However, with the development and spread of new technologies for recording from large groups of neurons, we have seen a resurgent interest in population-level thinking. Alongside new hardware, an explosion of new concepts and analyses have come to define the modern, population-level approach to neurophysiology (Yuste, 2015; Saxena and Cunningham, 2019; Jazayeri and Afraz, 2017; Shenoy and Kao, 2021; Vyas et al., 2020). Although advances in high-yield neural recordings were critical for the resurgence of these ideas, high-yield recordings are neither necessary nor sufficient to make a population neurophysiology paper. Instead, what defines the field is its object: the neural population. To a population neurophysiologist, neural recordings are not random samples of isolated units but are instead low-dimensional projections of the entire manifold of neural activity (Gallego et al., 2017; Jazayeri and Afraz, 2017). Here, we explain this and other fundamental ideas in population neurophysiology while illustrating why we believe this approach holds such promise for advancing our understanding of cognition.

**THE STATE SPACE**

For a single-unit neurophysiologist, the canonical analysis is a neuron’s peristimulus time histogram (PSTH). For a population neurophysiologist, it is a neural population’s state space diagram (Figure 1A). Instead of plotting the firing rate of one neuron against time, the state space diagram plots the activity of each neuron against one or more other neurons. At every moment in time, the population is at some neural state: it occupies some point in neuron-dimensional space or, identically, produces some vector of firing rates across recorded neurons. Time is a function that links neural states; it turns sequences of neural states (or sets of PSTHs) into trajectories through the state space (Figure 1B). (Trajectories have interesting implications for thinking about how the brain computes. We return to them in Dynamics.) Especially in cognitive studies, neural states may be called representations, a somewhat contentious term meant to highlight the correspondence between a neural state and some percept, memoranda, computation, or behavior (see Box 1).

Recasting population activity as a neural state can suggest new hypotheses. As vectors in neuron-dimensional space, neural states both point in some direction and have some magnitude (Figure 1A). Because the direction of a neural state vector is related to the pattern of activity across neurons, it is probably unsurprising that state vector direction encodes object identity in the inferotemporal (IT) cortex (Jaegle et al., 2019b; Chang and Tsao, 2017). However, it may be more surprising that this second feature—neural state magnitude—also matters: it predicts how well objects will be remembered later (Jaegle et al., 2019a, 2019b). When we pause to realize that magnitude is essentially...
a sum of activity across neurons, this may seem like an obvious analysis, or one that just recapitulates the cell-averaged PSTH. However, this is a sum across all neurons—regardless of their tuning properties—and it is taken without normalization. For a single-neuron neurophysiologist, it is essential to normalize the dynamic range of neurons before averaging—this is the only way to produce a clear picture of the response profile of an average neuron. However, for a population neurophysiologist, more concerned with the holistic population than the average neuron, differences in the dynamic range between neurons are a critical part of the signal. Because spikes are energetically costly and large-magnitude neural states require more spikes, understanding this aspect of the population code could be particularly interesting in cognitive domains where energetic efficiency is a concern, like in classic, energetically costly conflict signals (Ebitz et al., 2020a; Ebitz and Platt, 2015), or in understanding why we follow simple rules in lieu of less efficient but more flexible decision-making processes (Cohen et al., 2021; Ebitz et al., 2020b).

Because the state space gives us a spatial view of neural activity, it makes it natural to start to think about spatial relationships between different neural states: to reason about distances (Figure 1A). There are many ways to measure the distance between neural states, including the Euclidean distance, the angle between state space vectors (which decorrelates distance from differences in magnitude), and the Mahalanobis distance (which accounts for the covariance structure between neurons). (See Walther et al., 2016, for discussion of distance measures in neural data analysis.) Distance measures have many applications, including for reasoning about how neural states evolve over time. For example, in a slowly changing environment, we might expect neural states to also change slowly over time. However, there can be sudden jumps in neural states across pairs of trials (Karlsson et al., 2012; Durstewitz et al., 2010; Bartolo et al., 2020; Ebitz et al., 2018; Malagon-Vina et al., 2018). These jumps could reflect cognitive or behavioral discontinuities—sudden changes in our beliefs or policies—and are thus difficult to reconcile with simple cognitive models of sequential decision-making, in which information is integrated slowly across multiple trials and behavior changes gradually, rather than suddenly (Bartolo and Averbeck, 2020). Instead, jumps may better resonate with hierarchically structured models, in which inferences are being made at the level of policies, rather than actions (Collins and Koechlin, 2012; Ebitz et al., 2019, 2020b; Eckstein and Collins, 2020).

Measures of distance can also be combined with clustering algorithms from machine learning to examine the similarity between neural responses across task conditions (Kriegeskorte et al., 2008; Hunt et al., 2018), characterize hierarchical relationships between different pieces of information (Kiani et al., 2007; Reber et al., 2019), or examine the variance within or between task conditions—a factor that changes with learning (Thorn et al., 2010) and across goal or belief states (Ebitz et al., 2018). Perhaps the most systematized approach for reasoning about spatial relationships between neural states is representational similarity analysis (RSA), a method first developed in neuroimaging that has been extensively reviewed in that context (Kriegeskorte et al., 2008; Diedrichsen and Kriegeskorte, 2017). In neurophysiology, RSA has been used to ask how semantic knowledge shapes mnemonic representations (Reber et al., 2019), to compare the computations involved in decision-making across regions (Hunt et al., 2018), to characterize the stability
of representational structure over working memory (Spaak et al., 2017), to examine how rules shape information processing across regions (Ebitz et al., 2020b), and to contrast the contributions of different brain structures to encoding task dimensions (Keene et al., 2016). In fMRI, RSA is often interpreted as a kind of neural recapitulation of perceptual or psychological similarity. However, we still lack whole-brain resolution, so neurophysiologists should be cautious about these kinds of interpretations, because different representational structures may be found in different brain regions (Hunt et al., 2018; Keene et al., 2016).

Neural state space diagrams can have as many axes as there are recorded neurons. However, neural activity often only varies along a smaller number of directions in the state space, known as dimensions (Gao and Ganguli, 2015; Gao et al., 2017; Low et al., 2018; Gallego et al., 2017, 2018; Chaudhuri et al., 2019; Lehky et al., 2014). Because activities of different neurons are correlated with one another (Cohen and Kohn, 2011), every neuron does not make an independent contribution to the population (Umakantha et al., 2020). It is generally a combination of neurons that drives most variability in neural activity. We take advantage of this redundancy by using dimensionality reduction to compress neuron-dimensional state spaces onto a smaller number of axes (Cunningham and Yu, 2014). There are many ways to choose these axes, including principal-component analysis (PCA) (Figure 1C). Other dimensionality reduction methods, developed specifically for neurophysiology data, can take advantage of the temporal structure of these data and even solve other problems, like combining (stitching) nonsimultaneously recorded datasets (Box 2) (Yu et al., 2009; Pandarinath et al., 2018; Kobak et al., 2016; Williams et al., 2018). Because it is hard for humans to reason about higher-dimensional spaces, dimensionality reduction is helpful for exploratory analyses, but it can also be an important part of data processing pipelines because it can reduce the amount of noise in the data.

Is there always a one-to-one correspondence between a neural state and a particular cognitive state? There is some evidence that multiple patterns of activity could implement the same function at different moments in time (Malagon-Vina et al., 2018; Rule et al., 2020; Driscoll et al., 2017; Ebitz et al., 2018). This idea is known as multiple realizability in philosophy of mind. For example, in uncertain environments, decision-makers often pass through some period of exploration between longer periods of following some rule or policy (Wilson et al., 2021; Ebitz et al., 2018, 2019). Exploration produces the kinds of sudden jumps in neural activity we introduced with the concept of distances, but it also disrupts long-term autocorrelations between neural states and promotes new learning. The pattern of activity that implements a policy after exploration is not the same as the pattern that existed before it, even when subjects are just returning to an old policy. The ability to implement the same policy via slightly different neural states could offer some benefits in nonstationary environments (Ajemian et al., 2013; Chambers and Rumpel, 2017), but it also implies that the neural state spaces may have many sloppy dimensions—dimensions along which neural activity can vary without affecting cognition and behavior—and a
Box 2. Single-trial resolution

One draw of the population approach is the ability to decode from neural activity finer timescales than is possible with single neurons: the ability to take advantage of the lateral power we can achieve by combining across neurons, as opposed to the vertical power we would achieve by combining across trials (Stokes and Spaak, 2016). Population-level analyses have better temporal resolution than single-neuron analyses, because adding neurons to an analysis can add new dimensions (Stokes and Spaak, 2016; Leavitt et al., 2017; Umakantha et al., 2020). With more dimensions, we can increase the precision with which we can decode continuous information (because we can triangulate decoding across differently tuned neurons) or the reliability with which we can classify trials into their respective conditions (because adding dimensions improves the signal-to-noise ratio) (Figures 2G–2I).

Neurons do not have to be recorded simultaneously to take advantage of lateral power. Several prominent population neurophysiology studies were based on nonsimultaneously recorded neurons, which were later combined into pseudopopulations (Churchland et al., 2012; Machens et al., 2010; Mante et al., 2013; Meyers et al., 2009). A pseudopopulation is an estimate of what the population response might have been if neurons were recorded simultaneously. It is constructed through bootstrapping: sampling firing rates randomly with replacement from nonsimultaneously recorded neurons into a new collection of pseudotrials. Although the pseudopopulation approach disrupts the true correlation structure between neurons and is ill suited to studying events that occur at random times within trials or on a random subset of trials, it can deliver single-trial insight into neural computations. It thus remains a powerful way to perform population-level analyses of older datasets or in regions that are not yet amenable to high-yield recording technologies.

Some questions we can only answer with truly simultaneous, high-yield recording. The best example within cognitive neuroscience comes from perceptual decision-making. Perceptual decisions about time-varying signals are typically modeled as the slow integration of evidence in favor of one decision or another (Gold and Shadlen, 2007). Neural activity mirrors this hypothesized integration process: the firing rates of lateral intraparietal neurons ramp up slowly before the decision (Shadlen and Newsome, 2001; Gold and Shadlen, 2007). However, because single neurons are averaged over trials, ramping could also be the result of single neurons or the population shifting from some undecided, uncommitted state to a decided state at some random time point on each trial (Latimer et al., 2015; Churchland et al., 2011; Wong and Wang, 2006). Combining activity across trials would make activity appear to ramp up, even if the underlying generative process consisted of discrete steps at random times.

No clear consensus has emerged yet on the stepping-versus-ramping debate in perceptual decision-making (Zoltowski et al., 2019; Shadlen et al., 2016) and the truth is probably more complicated than either simplistic hypothesis (Zoltowski et al., 2019; Daniels et al., 2017). However, the new techniques developed over the course of this debate could help answer other important questions in cognitive neuroscience, like whether items held in working memory are maintained tonically or jumbled dynamically (Lundqvist et al., 2016; Miller et al., 2018). These methods could also help us uncover the causes and consequences of the dynamic alternation processes we see when multiple items compete for attention (Engel et al., 2016; Fiebelkorn and Kastner, 2019; Caruso et al., 2018) or decision-making (Rich and Wallis, 2016) or help us understand neural computations that may not be so tightly aligned with task events (Jones et al., 2007; Sadacca et al., 2016; Morcos and Harvey, 2016). We return to some of these ideas when we discuss dynamics.
dimensionalities might be better for different computations (Figure 2B). Theoretical work suggests that the dimensionality of neural activity could be yoked to the dimensionality of the task (Gao et al., 2017; Gao and Ganguli, 2015). Although it is a bit harder to think about tasks geometrically, they can, like manifolds, have an intrinsic dimensionality. If we choose to move left or right based on a single piece of information (like the direction of moving dots), we only need a 1-dimensional manifold to represent the evidence for left or right. If it matters that we decide at a particular time, we may need a second dimension to keep track of time, but the critical point is that laboratory tasks often have low dimensionality.

Although the number of dimensions is generally obvious in motor tasks, this is not always true of cognitive tasks (Akrami et al., 2018; Chen et al., 2021; Constantinople et al., 2019; Filipowicz et al., 2020; Glaze et al., 2018). For example, mice, when given an essentially 1-dimensional cognitive task (choosing between 2 images), made decisions that depended on irrelevant information (like image locations, past rewards, and past choices) (Chen et al., 2021). Behavior had many more dimensions than the task. This is neither specific to that task (Akrami et al., 2018; Constantinople et al., 2019) nor some artifact of the difficulty of designing good tasks for rodents. Humans also consistently overestimate the dimensionality of cognitive tasks (Glaze et al., 2018; Filipowicz et al., 2020). We argue that the brain is generally also engaged in task-irrelevant processing, and extra-task information would increase manifold dimensionality beyond the dimensionality of the task. Throughout the brain, we find signals that fluctuate with arousal (Ebitz and Platt, 2015; Vinck et al., 2015; Engel et al., 2016; Engel and Steinmetz, 2019) or irrelevant movements (Musall et al., 2019; Vinck et al., 2015). Indeed, there is some empirical evidence that manifolds can have higher dimensionality than the task (Low et al., 2018), although more work is needed to understand how the dimensionality of neural manifolds is related to the dimensionality of tasks and/or behavior.

Some theoretical work suggests that there may be computational benefits to higher-dimensional manifolds (Bernardi et al., 2020). With a higher-dimensional manifold, we have more flexibility in how we can decode information from neural activity. As dimensionality increases, new decoding strategies become possible that group increasingly arbitrary collections of neural states, meaning that a downstream structure could, in theory, use a decoding strategy that groups apples and penguins but excludes pears. If executive, prefrontal regions function partly to create this flexibility, then they should have higher dimensionality than other regions in the same task. However, few studies have compared neural manifolds across regions (although see Thura et al., 2020; Russo et al., 2020), and others might predict an opposing pattern: a progression from high-dimensional...
The neural manifold is a powerful concept for reasoning geometrically about neural activity, but we are only at the preci-
pice of understanding its implications for cognition. We do not yet know why a manifold will have a particular shape and/or dimensionality, how its geometric features are linked to cogni-
tion, or how they change across regions and why. Nor do we know what portion of the Manifold we are recovering in a few hours of recording in a specific task. Developing new methods to measure neural manifolds experimentally will be key.

**CODING DIMENSIONS**

One foundational idea in neurophysiology is that neurons are tuned for (i.e., respond differentially to) stimuli, motor responses, cognitive variables, or combinations thereof. Neural populations can also be tuned: neural states can covary with task information along specific directions in the neural state space. The directions that best correspond to (i.e., that are stiff with respect to) some stimuli, cognitive variables, or combinations thereof are known as coding dimensions (Figures 2C–2F). Coding dimensions are not just an aggregate property of tuned neurons: highly tuned neurons contribute to coding dimensions, but so do untuned neurons (Leavitt et al., 2017). Coding dimensions are thus an emergent property of a neural population.

Coding dimensions are closely linked to the problem of decoding from a neural population. We find coding dimensions by fitting a model to predict task information from neural data. Decoding from the model often involves a step that projects neural data onto the classification axis in the model. Critically, this step also has a geometric meaning: we are using the coding dimension as a new axis for representing neural activity, much like we did with the principal components (PCs) in Figure 1. However, here we are representing a neural state according to how A-like or B-like the population is, rather than where it falls along one PC or another. When used in this geometric sense, this process is known as targeted dimensionality reduction, because we are reducing the dimensionality of the population to only consider the axes along which it encodes some piece or pieces of information (Cunningham and Yu, 2014; Mante et al., 2013).

Coding dimensions and targeted dimensionality reduction are powerful tools for linking neurometric and psychometric func-
tions (Figures 2D–2F). For example, targeted dimensionality reduction was used to obtain, for the first time, a trial-by-trial measure of attentional modulations in visual area V4 (Cohen and Maunsell, 2010). Variability in each trials’ projection along the attention coding dimension predicted accuracy and response time in detecting a subtle stimulus change on single tri-
als. Targeted dimensionality reduction can also be used to ask whether trial-by-trial information about one variable (like the value of an option) predicts an entirely different variable (like the choice an animal will make) (McGinty and Lupkin, 2021)—linking neural computations to the behavior thought to depend on those computations. As the number of simultaneously collected neurons has continued to grow, so has the temporal precision we can achieve with these methods (Figures 2G–2I; Box 2). One recent study was even able to identify momentary fluctuations in decision variables in real time (Peixoto et al., 2021). This study is important not only because it set a new bar for the temporal precision of decoding but also because of its unique, closed-loop approach. By triggering pulses of infor-
mation according to the instantaneous value of the coding dimension projection, this study was able to show that these sig-
nals are not simply some statistical trick but are instead mean-
ingful indices of ongoing computations.

Coding dimensions need not be static, either within (Spaak et al., 2017; Stokes et al., 2013; Stokes, 2015; Lin et al., 2021; Kimmel et al., 2020; Cavanagh et al., 2018) or across trials (Mal-
agon-Vina et al., 2018; Rule et al., 2020; Driscoll et al., 2017; Ebizt et al., 2018). Across trials, we have already noted that mul-
tiple realizations of the same goal or belief state may be imple-
mented by slightly different patterns of activity (Malagon-Vina et al., 2018; Rule et al., 2020; Driscoll et al., 2017; Ebizt et al., 2018). However, coding dimensions can also be dynamic within trials. For example, in working memory, sustained activity in delay neurons was once thought to be responsible for maintain-
ing items in working memory. However, if delay neurons were solely responsible for maintaining memoranda, then population coding would be some consistent linear combination of delay neurons; it would not change over time. However, coding dimensions often change over the course of a memory period (Spaak et al., 2017; Cavanagh et al., 2018; Stokes, 2015). Dynamic co-
des could emerge from the need to keep track of time or reflect a transformation of sensory information into motor preparatory signals. However, they are not inevitable: other studies report
stable population codes within similar task epochs (Kimmel et al., 2020; Murray et al., 2017).

What can we learn from coding dimension geometry? Some researchers are working to understand geometric features of coding dimensions, like their curvature (Sohn et al., 2019; Okazawa et al., 2021; Thura et al., 2020). In this context, the term manifold reappears, used now to refer to the (often nonlinear) shape of neural states that covary with some task variable. There is growing evidence that some linear variables may be encoded along a curved surface in the neural state space (Sohn et al., 2019; Okazawa et al., 2021). It remains unclear whether curvature facilitates a particular readout (Sohn et al., 2019), encodes some variable in its own right, or is an epiphenomenal consequence of constraints on firing rates (Okazawa et al., 2021). In any case, these results presage a future in which understanding the shape of coding dimensions could provide insight into cognitive functions.

Future studies may be able to use the relative geometry of different coding dimensions to understand cognition. For example, looking at a recurrent neural network (RNN) trained to perform various tasks, a recent study found striking alignment between coding dimensions that corresponded to specific cognitive functions across tasks (Yang et al., 2019). Population activity was displaced in the same direction in the neural state space whenever the task required an item to be held in working memory, for example. This implies that working memory was executed by a particular coding dimension that was reused across tasks. Experimental work is needed to determine whether the brain can compose either coding dimensions or cognitive functions, but this study illustrates two important points. First, coding dimensions could be the key to answering big questions in cognitive neuroscience, like whether cognitive functions are composable or even unitary. Second, theoretical work is critical for generating new hypotheses and neural population analyses.

**SUBSPACES**

To make sense of heterogeneity across individual neurons, we have traditionally grouped them into classes. At the population level, we analogously identify subspaces. We can think of a subspace as a lower-dimensional projection of the neural state space: a restricted number of dimensions within the larger state space (Cunningham and Yu, 2014). For example, each neuron’s response could be considered a subspace of the population, or a subspace could be composed of some weighted combination of neuronal firing rates across the population. Although the term subspace is sometimes used loosely in neuroscience, a set of dimensions only forms a proper, mathematical subspace if the dimensions are orthogonal to one another. Caution is important here, because projecting neural activity into nonorthogonal axes can warp the relationships between neural states in unexpected ways.

Often, we are interested in subspaces that encode a piece of information or perform some function. Imagine a neural population that is mixed selective for stimulus color and other information, like location (Rigotti et al., 2013; Raposo et al., 2014). The color subspace of this population would be the portion of the neural space spanned by the color-coding dimensions (e.g., red-, green-, and blue-coding axes) and the location subspace would be the portion spanned by location-coding dimensions. These subspaces can be thought of as higher-dimensional generalizations of the coding dimension: neural activity is only stiff with respect to the encoded variable within the subspace. Recent work has looked at the neural subspaces responsible for processing different task dimensions, particularly in perceptual decision-making (Mante et al., 2013; Aoi et al., 2020) and value-based decision-making (Ebitz et al., 2020b). The latter study compared color- and shape-subspace projections while monkeys followed simple color- and shape-based sensorimotor rules (like “choose blue things” or “choose triangles”). It found that information coding in different subspaces was not fixed but instead was gated by their relevance to the current goal of the animal. This could explain why rule-based decisions were also more energetically efficient than other types of decisions in this study; perhaps rules make efficient use of limited neural resources, because using them allows us to collapse the goal-irrelevant dimensions of the neural population code.

One special kind of subspace is a nullspace: a subspace that is not associated with some function or piece of information (Kaufman et al., 2014). This term was first used to refer to the portion of the neural state space that bears no relationship to motor effectors and thus is likely more involved in the preparatory or cognitive aspects of motor control (Kaufman et al., 2014). We could also imagine a nullspace in which neural activity is not related to stimulus color, location, or the choice the monkey will make (i.e., the dimensions of neural activity that are sloppy with respect to these variables). There is a trivial way to create a nullspace (or any other kind of subspace): have a particular function performed by some segregated group of neurons within the population. If motor effectors were only meaningfully coupled to the spinal-projecting neurons in motor cortex, for example, then it would be trivially true that we would find a motor effector subspace composed only of these neurons. It would also be trivially true that we would have a motor effector nullspace composed of all other neurons. However, Kaufman et al. (2014) found that this was not the case. The nullspace in this paper was composed of a weighted combinations of neurons, rather than a segregated population. This result strongly resonates with the idea that mixed selectivity is a ubiquitous property of single neurons (Rigotti et al., 2013; Raposo, Kaufman, and Churchland, 2014)—an idea that implies that it is the pattern of activity across a population of neurons, rather than a segregated group, that implements some function or encodes some piece of information.

The term subspace can also be used in a way that draws inspiration from neural manifolds, rather than coding dimensions. Not all neural dimensions are needed during each epoch of a task. During a task epoch that does not require working memory, for example, neural dimensions that are responsible for working memory should not be occupied. Thus, by identifying the portions of the neural manifold that are occupied across epochs—the subspaces that define each epoch—we can gain insight into the relationship between the cognitive operations that are occurring (Elsayed et al., 2016). For example, one decision-making study compared the neural subspaces across
epochs when subjects viewed two offer cues in sequence (Yoo and Hayden, 2020). If the neural populations simply encoded offer values, we would expect these subspaces to be identical. However, the second offer was encoded in a different subspace from the first so that the evaluation and comparison steps were functionally segregated.

We have seen that subspaces can be defined based on their relationship to task variables, epochs, and motor effectors (Kaufman et al., 2014), but they can also be defined by their relationship to other brain regions (Semedo et al., 2019). Here, instead of using regression to identify a coding dimension that maps neural activity to some task dimension, researchers identify dimensions in one neural population that explain variability in a second population. This approach has the potential to be incredibly powerful, because it can isolate the information shared by two regions in the communication subspace (Semedo et al., 2019), the private information within each region’s nullspace, and determine how this information changes over time (Perich et al., 2020). However, these methods have the same limitations as all correlational methods: they assume some directionality but cannot uniquely determine how information flows between regions or whether it is inherited from some third region that influences both.

DYNAMICS

Much of the work we have talked about so far assumes that neural activity can be summarized by a single neural state or a single point in time. However, neural activity clearly evolves over time and across neurons in important ways. A single-neuron neurophysiologist might examine these patterns by scrolling through the PSTHs of the individual neurons or averaging them together. At the population level, we simultaneously consider a collection of PSTHs by examining neural trajectories: the paths that activity traces through the neural state space (Figure 1B). Why do neural trajectories evolve as they do? One influential idea is that trajectories are shaped by hidden network-level forces called dynamics (Figure 3A). If we imagine neural activity as a ball rolling across a landscape, the trajectory is the path traced by the ball, but the dynamics are the forces that shape its path: the landscape itself. We generally cannot measure dynamics, because they are a complex function of features we cannot resolve experimentally: the networks’ wiring, inputs, and activation functions (although see Genkin et al., 2020). However, we can make inferences about dynamics by looking at the frequency of different neural states, examining the shape of neural trajectories, comparing data with computational models, or looking at how the network responds to perturbations. There are many excellent reviews on population dynamics (Brody et al., 2003; Vyas et al., 2020; Sussillo, 2014; Chaudhuri and Fiete, 2016; Shenoy and Kao, 2021; Yuste, 2015; Ju and Bassett, 2020), so we focus on their links to cognitive processes here.

Perhaps the most fundamental concept in population dynamics is the attractor. An attractor is a valley in the landscape, a neural state (or collection thereof) toward which nearby activity will evolve. Many neural network architectures naturally produce multiple stable, fixed-point attractors, and theoretical work has long implicated these in cognition (Moreno-Bote et al., 2007; Hopfield, 1982; Miller, 2016; Wong and Wang, 2006). For example, a Hopfield network (an early RNN) shows how a system of attractors can solve important problems in associative memory (Hopfield, 1982). Because they pull nearby patterns of activity toward a stable state, attractors are ideally suited to solve cognitive problems like categorizing an input or holding an item stably in memory (Miller, 2016).

When a network includes multiple attractors, we can see evidence of multi-stability in its activity (Miller, 2016): the network transitions between essentially discrete neural states (Figure 3B). Theoretical work suggests that this is a natural consequence of clustered connections between neurons (Litwin-Kumar and Doiron, 2012). In neurophysiology data, multi-stable patterns often appear when multiple objects are in competition, be it for visual attention (Engel et al., 2016; Fiebelkorn and Kastner, 2019) or auditory attention (Caruso et al., 2018), or in value-based decision-making (Rich and Wallis, 2016; Hayden and Moreno-Bote, 2018). For example, one visual attention study reported multi-stable population states within cortical columns in visual cortex (Engel et al., 2016): neurons selective for one stimulus were either on (highly active) or off (largely inactive), a striking phenomenon observable only at the population level. Attention regulated the duration of on states, which in turn predicted the detection of subtle stimulus changes and may even explain some single-neuron correlates of selective attention. In this study, the transitions between states were stochastic, but attention can also be reoriented rhythmically (Dugué et al., 2016; Fiebelkorn and Kastner, 2019) and future work is
needed to understand whether transitions between multi-stable attractors are driven rhythmically or stochastically.

Sequences of multi-stable attractors could implement computations (Miller, 2016; Vyas et al., 2020; Sussillo, 2014; Schittler Neves and Timme, 2012). For example, the classic attractor network model of decision-making evolves from an undecided state to one of several possible absorbing, decided states over time, with evidence (Wong and Wang, 2006; Wang, 2008). This style of model predicts that perturbing neural activity should have a larger effect early in the process, before the system has settled into a decided state, than it does later near the attractor, where stabilizing forces are greater. Indeed, this is seen experimentally with optogenetic perturbations in rodent decision-making regions (Kopec et al., 2015).

These models are highly simplified. They can be implemented in a network containing only 2-3 neuronal pools, and not all decision-making is best made by ballistic, winner-take-all dynamics. In a more naturalistic, consummatory decision-making paradigm, animals transitioned between sequences of 4-5 discrete neural states as they evaluated and decided whether to accept or reject a proffered reward (Sadacca et al., 2016). Other more complex neural network models cast perceptual decision-making as the product of a line attractor (Sussillo, 2014; Mante et al., 2013), meaning a set of multi-stable attractors that neural activity is driven along by each new piece of evidence. Different regions may also implement different dynamics in the same task. For example, in rats performing an evidence accumulation task, researchers found evidence of graded, line-attractor-like representations in posterior cortex, but ballistic, winner-take-all dynamics in frontal cortex (Hanks et al., 2015), despite nearly identical averaged PSTHs in the two regions.

We have considered only stable, fixed-point attractors, but attractors are not necessarily stable in all dimensions, especially in high-dimensional systems. Activity may be pulled into an attractor from one direction only to be pushed away, toward another attractor, along another direction. Collections of this type of partially stable attractor are known as heteroclinic channels, and the trajectories they produce are known as heteroclinic cycles or orbits (Schittler Neves and Timme, 2012; Rabinovich et al., 2001, 2008), although they can have any shape. In neuroscience, the term dynamic attractor is perhaps more common (Laje and Buonomano, 2013). Dynamic attractors can produce complex, continuously evolving trajectories through the neural state space and are, in theory, capable of implementing any arbitrary computation (Schittler Neves and Timme, 2012).

OPEN QUESTIONS

Population neurophysiology has its own object of study, characteristic set of methods, and suite of key concepts that give us new ways to reason about how neurons behave collectively, rather than as individuals. We have introduced 5 of these concepts here: (1) the neural states that provide a snapshot of a pattern of activity across the population, (2) the manifold that encompasses the neural states that are possible (Manifold) or at least observed (manifold), (3) the coding dimensions and (4) the subspaces that link neural states to behavior and cognition, and (5) the dynamics that map activity from neural state to neural state, guiding how trajectories evolve through time and across the state space. These concepts have links to ideas and analyses from the single-neuron approach—links that we have worked to highlight here—but despite our best efforts, these mappings are not perfect. There is something new about population-level thinking that simply cannot be understood as a composition of single-neuron concepts, just as the population itself cannot be understood as a composition of single neurons. To us, these are signs that population neurophysiology is coalescing into a new field. However, important conceptual and methodological questions remain.

Conceptually, we should acknowledge that the neural population doctrine has a weakness that is not shared with the single-neuron doctrine. The limits of a neuron are obvious—it has cell walls—but what are the limits of a population? Are its boundaries the set of recorded neurons? The tissue surrounding the electrodes? The edges of the Broadman area? The skull? Advances in neural recording technologies may move us toward this broadest notion of a neural population, but for now, the term is ambiguous. It is not always immediately clear whether a paper shares our notion of population or when the term population is distinct from related terms, like neuronal ensembles. It may not be necessary to have a formal definition, provided practitioners of the population approach are able to understand one another, but the lack of a definition does present opportunities for confusion.

Methodologically, it is not clear that we have fully come to terms with some implications of the population doctrine. If our goal is to understand how neural populations behave collectively, is spike sorting still necessary or does it become irrelevant? Spike sorting is an imperfect process, and despite concerted efforts to automate it, the process remains incredibly time consuming. This bottleneck will only grow as neuronal yields continue to accelerate. Although state-of-the-art algorithms perform well, they require experimentalists to sacrifice the spatial extent of recordings for dense, overlapping coverage of a smaller regions (Rossant et al., 2016). However, isolating individual neurons is probably not critical for every population-level analysis (Trautmann et al., 2019). Developing a better understanding of when isolated cells matter and when they do not could help us move to a model in which neurons are isolated only when scientifically necessary.

We are just starting to explore how this new generation of population-level analyses may connect with other population-level phenomena, like neuronal correlations and field potentials. It is clear that correlations between neurons have substantial, structuring effects on population-level representations (Uma kantha et al., 2020; Elsayed and Cunningham, 2017). In our view, continued work on these correlations will be an essential bridge between single-neuron and population-level accounts in the future. Local field potentials are emergent, population-level phenomena in their own right. Although we have not yet been explicit on this point, oscillations across populations of neurons can be used to index population computations in ways that can either parallel spiking results or deliver important new insights (Hunt et al., 2015; Chaudhuri et al., 2019; Gallego-Carracedo et al., 2021; Hall et al., 2014; Lundqvist et al., 2016; Smith et al., 2019; Widge et al., 2019). If population neurophysiology...
continues to evolve toward a broader notion of the limits of a population, perhaps we will see more researchers embrace the insights that are only possible with broad-scale electrophysiological measures like local field potentials, electrocorticography, electroencephalography, and/or magnetoencephalography.

**Conclusions**

Some scholars have argued that it is time to replace the single-neuron doctrine with the neural population doctrine (Saxena and Cunningham, 2019). We are sympathetic to this view. So was Canadian neuroscientist Donald Hebb, who argued that the assembly—a collection of cells wired together—should be considered the fundamental unit of computation, because it produces reverberating behavior that cannot be understood by isolating each neuron (Hebb, 1949). Theoretical neuroscientists have long worked at the level of the population, extrapolating insights from single neurons into population-level computational models (Laje and Buonomano, 2013; Hopfield, 1982; Wong and Wang, 2006; Bernardi et al., 2020; Moreno-Bote et al., 2007; Chaudhuri and Fiete, 2016; Brody et al., 2003; Miller, 2016; Sussillo, 2014; Yuste, 2015; Litwin-Kumar and Doiron, 2012). Clearly, this view is not new. What is new is the ability to readily record from large numbers of neurons, the development of new tools to analyze population activity, and a suite of concepts that constitute a systematic new framework for reasoning about neural activity at the population level.

Some have argued that population-level phenomena must be emergent, that is, exist only at the population level, to be interesting (Saxena and Cunningham, 2019; Pillow and Aoi, 2017; Elsayed and Cunningham, 2017). Focusing on emergent phenomena may be the most straightforward way to overcome the occasional misconception that population neurophysiology is trivially packaging of what we already knew, only with more neurons. However, it is not necessary for a phenomenon to exist only at one level to be relevant to that level. Even if a population-level result could be entirely predicted from the properties of single neurons, did anyone make the prediction? Does the population-level description give us a new hypothesis to test, a more compact description, or a surprising new view? We must remain cautious in our armchair intuitions about how single-neuron results will map to the population level. We almost certainly could have intuited the existence of coding dimensions may be the most straightforward way to overcome the occasional misconception that population neurophysiology is trivially packaging of what we already knew, only with more neurons. However, it is not necessary for a phenomenon to exist only at one level to be relevant to that level. Even if a population-level result could be entirely predicted from the properties of single neurons, did anyone make the prediction? Does the population-level description give us a new hypothesis to test, a more compact description, or a surprising new view? We must remain cautious in our armchair intuitions about how single-neuron results will map to the population level. We almost certainly could have intuited the existence of coding dimensions from single-neuron tuning, but our intuition about how tuned neurons contribute to coding dimensions would have been wrong (Leavitt et al., 2017), we probably would not correctly predict coding dimension geometry (Okazawa et al., 2021), and we may not have even considered that coding dimensions may be dynamic (Stokes et al., 2013). Translating ideas in new frameworks is rarely trivial, even if it may take a few years to know whether a population-level explanation is novel or just a critical bridge between increasingly disparate levels of analysis.

**ACKNOWLEDGMENTS**

This work was supported by the Natural Sciences and Engineering Research Council (Discovery Grant RGPIN-2020-05577 to R.B.E.), the Fonds de la recherche en santé du Québec (Junior 1 Salary Award 284309 to R.B.E.), the Brain and Behavior Research Foundation (Young Investigator Grant 27298 to R.B.E.), and the National Institutes of Health (NIDA 038615 and NINDS 118366 to B.Y.H.). We thank Katarzyna Jurewicz, Alex Herman, Vince McClinty, Justin Fine, Pouya Bashivan, Suresh Krishna, and three anonymous reviewers for helpful comments on previous versions of this manuscript.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**REFERENCES**

Ajemian, R., D’Ausilio, A., Moorman, H., and Bizzi, E. (2013). A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits. Proc. Natl. Acad. Sci. USA 110, E5075–E5087.

Akrani, A., Kopec, C.D., Diamond, M.E., and Brody, C.D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. Nature 554, 368–372.

Aoi, M.C., Mante, V., and Pillow, J.W. (2020). Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. Nat. Neurosci. 23, 1410–1420.

Baker, B., Lansdell, B., and Kording, K. (2021). A Philosophical Understanding of Representation for Neuroscience. arXiv, 2102.06592 https://arxiv.org/abs/2102.06592.

Barack, D.L., and Krakauer, J.W. (2021). Two views on the cognitive brain. Nat. Rev. Neurosci. 22, 359–371.

Bartolo, R., and Averbeck, B.B. (2020). Prefrontal cortex predicts state switches during reversal learning. Neuron106, 1044–1054.

Bartolo, R., Saunders, R.C., Mitz, A.R., and Averbeck, B.B. (2020). Dimensionality, information and learning in prefrontal cortex. PLoS Comput. Biol. 16, e1007514.

Bernardi, S., Benna, M.K., Rigotti, M., Munuera, J., Fusi, S., and Salzman, C.D. (2020). The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. Cell 183, 954–967.e21.

Brette, R. (2019). Is Coding a Relevant Metaphor for the Brain? Behav. Brain Sci. 42, e215.

Brody, C.D., Romo, R., and Kepecs, A. (2003). Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. Curr. Opin. Neurobiol. 13, 204–211.

Canuso, V.C., Mohl, J.T., Glynn, C., Lee, J., Willett, S.M., Zaman, A., Eibihara, A.F., Estrada, R., Freewald, W.A., Tokdar, S.T., and Groh, J.M. (2018). Single neurons may encode simultaneous stimuli by switching between activity patterns. Nat. Commun. 9, 2715.

Cavanagh, S.E., Towers, J.P., Wallis, J.D., Hunt, L.T., and Kennerley, S.W. (2018). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. Nat. Commun. 9, 3498.

Chambers, A.R., and Rumpel, S. (2017). A stable brain from unstable components: Emerging concepts and implications for neural computation. Neuroscience 357, 172–184.

Chandrasekaran, C., Soldado-Magraner, J., Peixoto, D., Newsome, W.T., Shenoy, K.V., and Sahani, M. (2018). Brittleness in Model Selection Analysis of Single Neuron Firing Rates. BioRxiv. https://doi.org/10.1101/430710.

Chang, L., and Tsao, D.Y. (2017). The Code for Facial Identity in the Primate Brain. Cell 169, 1013–1028.e14.

Chaudhuri, R., and Fiete, I. (2016). Computational principles of memory. Nat. Neurosci. 19, 394–403.

Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., and Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nat. Neurosci. 22, 1512–1520.
Chen, C.S., Ebitz, R.B., Bindas, S.R., Redish, A.D., Hayden, B.Y., and Griswom, N.M. (2021). Divergent strategies for learning in males and females. Curr. Biol. 31, 39–50.e4.

Churchland, A.K., Kiani, R., Chauhuri, R., Wang, X.J., Pouget, A., and Shadlen, M.N. (2011). Variance as a signature of neural computations during decision making. Neuron 69, 818–831.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature 487, 51–56.

Churchland, P.S., and Sejnowski, T.J. (1990). Neural Representation and Neural Computation. Philos. Perspect. 4, 343–382.

Cisek, P. (2006). Preparing for speed. Focus on “Preparatory activity in premotor cortex reflects the speed of the upcoming reach”. J. Neurophysiol. 96, 2842–2843.

Cohen, M.R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. Nat. Neurosci. 14, 811–819.

Cohen, M.R., and Maunsell, J.H.P. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. J. Neurosci. 30, 15241–15253.

Cohen, Y., Schneidman, E., and Paz, R. (2021). The geometry of neuronal representations during rule learning reveals complementary roles of cingulate cortex and putamen. Neuron 109, 839–851.e9.

Collins, A., and Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. PLoS Biol. 10, e1001293.

Constantinople, C.M., Piet, A.T., Bibawi, P., Akrami, A., Kopec, C., and Brody, C.D. (2019). Lateral orbitofrontal cortex promotes trial-by-trial learning of risky, but not spatial, biases. eLife 8, e49744.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. 17, 1500–1509.

Daniels, B.C., Flack, J.C., and Krakauer, D.C. (2017). Dual Coding Theory Explains Bifasic Collective Computation in Neural Decision-Making. Front. Neurosci. 11, 313.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? Neuron 73, 415–434.

Diedrichsen, J., and Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity mapping. PLoS Comput. Biol. 13, e1005508.

Discoli, L.N., Pettit, N.L., Minderer, M., Chethi, S.N., and Harvey, C.D. (2017). Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. Cell 170, 986–999.e16.

Dugué, L., Roberts, M., and Carrasco, M. (2016). Attention Reorient(s) Periodically. Curr. Biol. 26, 1595–1601.

Ebitz, R.B., and Platt, M.L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. Neuron 85, 628–640.

Durstewitz, D., Vitoz, N.M., Floresco, S.B., and Seamans, J.K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron 6, 436–448.

Ebitz, R.B., Albarran, E., and Moore, T. (2018). Exploration Disrupts Choice-Predictive Signals and Alters Dynamics in Prefrontal Cortex. Neuron 97, 450–461.e9.

Ebitz, R.B., Sleezer, B.J., Jedema, H.P., Bradberry, C.W., and Hayden, B.Y. (2019). Tonic exploration governs both flexibility and lapses. PLoS Comput. Biol. 15, e1007457.

Ebitz, R.B., Smith, E.H., Horga, G., Schevon, C.A., Yates, M.J., McKhann, G.M., Botvinick, M.M., Sheth, S.A., and Hayden, B.Y. (2020a). Human Dorsal Anterior Cingulate Neurons Signal Conflict by Amplifying Task-Relevant Information. bioRxiv. https://doi.org/10.1101/2020.03.14.991745.

Ebitz, R.B., Tu, J.C., and Hayden, B.Y. (2020b). Rules warp feature encoding in decision-making circuits. PLoS Biol. 18, e3000951.

Eckstein, M.K., and Collins, A.G. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. Proc. Natl. Acad. Sci. USA 117, 29381–29389.

Elsayed, G.F., and Cunningham, J.P. (2017). Structure in neural population recordings: an expected byproduct of simpler phenomena? Nat. Neurosci. 20, 1310–1318.

Elsayed, G.F., Lara, A.H., Kaufman, M.T., Churchland, M.M., and Cunningham, J.P. (2016). Reorganization between preparatory and movement population responses in motor cortex. Nat. Commun. 7, 13239.

Engel, T.A., and Steinmetz, N.A. (2019). New perspectives on dimensionality and variability from large-scale cortical dynamics. Curr. Opin. Neurobiol. 58, 181–190.

Engel, T.A., Steinmetz, N.A., Gieselmann, M.A., Thiele, A., Moore, T., and Boahen, K. (2016). Selective modulation of cortical state during spatial attention. Science 354, 1140–1144.

Fetz, E.E. (1992). Are Movement Parameters Recognizably Coded in the Activity of Single Neurons? Behav. Brain Sci. 15, 679–690.

Fiebelkorn, I.C., and Kastner, S. (2019). A Rhythmic Theory of Attention. Trends Cogn. Sci. 23, 87–101.

Filipowicz, A.L.S., Glaze, C.M., Kable, J.W., and Gold, J.I. (2020). Pupil Diameter Encodes the Idiosyncratic, Cognitive Complexity of Belief Updating. Elife 9, e57972.

Fine, J.M., and Hayden, B.Y. (2021). The whole prefrontal cortex is premotor cortex. arXiv, 2106.04651 https://arxiv.org/abs/2106.04651.

Gallego, J.A., Perich, M.G., Miller, L.E., and Solla, S.A. (2017). Neural Manifolds for the Control of Movement. Neuron 94, 978–984.

Gallego, J.A., Perich, M.G., Naufel, S.N., Ethier, C., Solla, S.A., and Miller, L.E. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. Nat. Commun. 9, 4233.

Gallego, J.A., Perich, M.G., Chowdhury, R.H., Solla, S.A., and Miller, L.E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. Nat. Neurosci. 23, 260–270.

Gallego-Carracedo, C., Perich, M.G., Chowdhury, R.H., Miller, L.E., and Gallego, J.A. (2021). Local Field Potentials Reflect Cortical Population Dynamics in a Region-Specific and Frequency-Dependent Manner. BioRxiv. https://doi.org/10.1101/2021.05.31.446454.

Gao, P., and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. Curr. Opin. Neurobiol. 32, 148–155.

Gao, P., Trautmann, E., Yu, B., Santhanan, G., Ryu, S., Shenoy, K., and Ganguli, S. (2017). A Theory of Multineuronal Dimensionality, Dynamics and Measurement. BioRxiv. https://doi.org/10.1101/214262.

Genkin, M., Hughes, O., and Engel, T.A. (2020). Learning Non-Stationary Langevin Dynamics from Stochastic Observations of Latent Trajectories. arXiv, 2012.14944 https://arxiv.org/abs/2012.14944.

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

Glaze, C.M., Filipowicz, A.L.S., Kable, J.W., Balasubramanian, V., and Gold, J.I. (2018). A Bias-Variance Trade-off Governs Individual Differences in Online Learning in an Unpredictable Environment. Nat. Hum. Behav. 2, 213–224.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. 30, 555–574.

Hall, T.M., de Carvalho, F., and Jackson, A. (2014). A common structure underlies low-frequency cortical dynamics in movement, sleep, and sedation. Nature 515, 1185–1199.

Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., and Brody, C.D. (2013). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature 520, 220–223.

Hayden, B.Y., and Moreno-Bote, R. (2018). A neuronal theory of sequential economic choice. Brain Neurosci. Adv. 2, 2398212618766675.
Hayden, B.Y., and Niv, Y. (2021). The Case against Economic Values in the orbitofrontal cortex (or anywhere else in the Brain). Behav. Neurosci. 135, 192–201.

Hebb, D.O. (1949). The Organization of Behavior; a Neuropsychological Theory (Wiley).

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. Front. Hum. Neurosci. 3, 31.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79, 2554–2558.

Hunt, L.T., and Hayden, B.Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. Nat. Rev. Neurosci. 18, 172–182.

Hunt, L.T., Behrens, T.E., Hosokawa, T., Wallis, J.D., and Kennerley, S.W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. eLife 4, e11945.

Hunt, L.T., Malalasekera, W.M.N., deBerker, A.O., Miranda, B., Farmer, S.F., Behrens, T.E.J., and Kennerley, S.W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. Nat. Neurosci. 21, 1471–1481.

Jaegle, A., Mehrpour, V., Mohsenzadeh, Y., Meyer, T., Oliva, A., and Rust, N. (2019a). Population response magnitude variation in inferotemporal cortex predicts image memorability. eLife 8, e47596.

Jaegle, A., Mehrpour, V., and Rust, N. (2019b). Visual novelty, curiosity, and intrinsic reward in machine learning and the brain. Curr. Opin. Neurobiol. 58, 167–174.

Jazayeri, M., and Afnaz, A. (2017). Navigating the Neural Space in Search of the Neural Code. Neuron 93, 1003–1014.

Jones, L.M., Fontanini, A., Sadacca, B.F., Miller, P., and Katz, D.B. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. Proc. Natl. Acad. Sci. USA 104, 18772–18777.

Ju, H., and Bassett, D.S. (2020). Dynamic representations in networked neural systems. Nat. Neurosci. 23, 908–917.

Karlsson, M.P., Tervo, D.G., and Karpova, A.Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science 338 (6103), 123–123.

Kauffman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: permitting preparation without movement. Nat. Neurosci. 17, 440–448.

Keene, C.S., Blardon, J., McKenzie, S., Liu, C.D., O’Keefe, J., and Eichenbaum, H. (2016). Complementary Functional Organization of Neuronal Activity Patterns in the Perirhinal, Lateral Entorhinal, and Medial Entorhinal Cortices. J. Neurosci. 36, 3680–3675.

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

Kimmel, D.L., Elsayed, G.F., Cunningham, J.P., and Newsome, W.T. (2020). Value and choice as separable and stable representations in orbitofrontal cortex. Nat. Commun. 11, 3466.

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C.E., Kepets, A., Mainen, Z.F., Qi, X.L., Romo, R., Uchida, N., and Machena, C.K. (2016). Demixed principal component analysis of neural population data. eLife 5, e10889.

Kopec, C.D., Erlisch, J.C., Brunton, B.W., Deisseroth, K., and Brody, C.D. (2015). Cortical and Subcortical Contributions to Short-Term Memory for Orienting Movements. Neuron 88, 367–377.

Kosuth, J. (1965). “One and three chairs.” Wood chair, photograph of chair, and photograph of the dictionary definition of “chair” (Museum of Modern Art).

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.

Laje, R., and Buonomano, D.V. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. Nat. Neurosci. 16, 925–933.

Latimer, K.W., Yates, J.L., Meister, M.L., Huk, A.C., and Pillow, J.W. (2015). NEURONAL MODELING. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. Science 349, 184–187.

Leavitt, M.L., Pieper, F., Sachs, A.J., and Martínez-Trujillo, J.C. (2017). Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. Proc. Natl. Acad. Sci. USA 114, E2494–E2503.

Lehky, S.R., Kiani, R., Esteky, H., and Tanaka, K. (2014). Dimensionality of object representations in monkey inferotemporal cortex. Nat. Neurosci. 26, 2135–2162.

Lin, J.-Y., Stone, B.T., Herzog, L.E., Nanu, R., Mahmood, A., and Katz, D.B. (2021). The Function of Groups of Neurons Changes from Moment to Moment. Curr. Opin. Physiol. 20, 1–7.

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat. Neurosci. 15, 1498–1505.

Low, R.J., Lewallen, S., Aronov, D., Nevers, R., and Tank, D.W. (2018). Probing Variability in a Cognitive Map Using Manifold Inference from Neural Dynamics. BioRxiv. https://doi.org/10.1101/418939.

Lundqvist, M., Rose, J., Herman, P., Brinca, S.L., Buschman, T.J., and Miller, E.K. (2016). Gamma and Beta Bursts Underlie Working Memory. Neuron 90, 152–164.

Machens, C.K., Romo, R., and Brody, C.D. (2010). Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. J. Neurosci. 30, 350–360.

Malagon-Vina, H., Ciocchi, S., Passecker, J., Dorffner, G., and Klausberger, T. (2018). Fluid network dynamics in the prefrontal cortex during multiple strategy switching. Nat. Commun. 9, 309.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84.

McGinty, V.B., and Lupkin, S.M. (2021). Value Signals in Orbitofrontal Cortex Predict Economic Decisions on a Trial-to-Trial Basis. BioRxiv. https://doi.org/10.1101/2021.03.11.434452.

Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. J. Neurophysiol. 100, 1407–1419.

Michals, J.A., Dann, B., and Scherberger, H. (2016). Neural Population Dynamics during Reaching Are Better Explained by a Dynamical System than Representational Tuning. PLoS Comput. Biol. 12, e1005175.

Miller, E.K., Lundqvist, M., and Bastos, A.M. (2018). Working Memory 2.0. Neuron 100, 463–475.

Miller, P. (2016). Itinerary between Attractor States in Neural Systems. Curr. Opin. Neurobiol. 40, 14–22.

Morcos, A.S., and Harvey, C.D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. Nat. Neurosci. 19, 1672–1681.

Moreno-Bote, R., Rinzel, J., and Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. J. Neurophysiol. 98, 1125–1139.

Muller, T.H., Mars, R.B., Behrens, T.E., and O’Reilly, J.X. (2019). Control of entropy in neural models of environmental state. eLife 8, e39404.

Murray, J.D., Bernacchia, A., Roy, N.A., Constantinidis, C., Romo, R., and Wang, X.-J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proc. Natl. Acad. Sci. USA 114, 394–399.

Musall, L., Kuffa, M.T., Juavine, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci. 22, 1677–1686.

Oby, E.R., Golub, M.D., Hennig, J.A., Degenhart, A.D., Tyler-Kabara, E.C., Yu, B.M., Chase, S.M., and Batista, A.P. (2019). New neural activity patterns emerge with long-term learning. Proc. Natl. Acad. Sci. USA 116, 15210–15215.
Wilson, R.C., Bonawitz, E., Costa, V.D., and Ebitz, R.B. (2021). Balancing exploration and exploitation with information and randomization. Curr. Opin. Behav. Sci. 38, 49–56.

Wong, K.-F., and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. 26, 1314–1328.

Yang, G.R., Joglekar, M.R., Song, H.F., Newsome, W.T., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. Nat. Neurosci. 22, 297–306.

Yoo, S.B.M., and Hayden, B.Y. (2018). Economic Choice as an Untangling of Options into Actions. Neuron 99, 434–447.

Yoo, S.B.M., and Hayden, B.Y. (2020). The Transition from Evaluation to Selection Involves Neural Subspace Reorganization in Core Reward Regions. Neuron 105, 712–724.e4.

Yoo, S.B.M., Hayden, B.Y., and Pearson, J.M. (2021). Continuous decisions. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20190664.

Yu, B.M., Cunningham, J.P., Santhanan, G., Ryu, S.I., Shenoy, K.V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. J. Neurophysiol. 102, 614–635.

Yu and Hayden, B.Y. (2015). From the neuron doctrine to neural networks. Nat. Rev. Neurosci. 16, 487–497.

Zoltowski, D.M., Latimer, K.W., Yates, J.L., Huk, A.C., and Pillow, J.W. (2019). Discrete Stepping and Nonlinear Ramping Dynamics Underlie Spiking Responses of LIP Neurons during Decision-Making. Neuron 102, 1249–1258.e10.