

Causes and consequences of representational drift

Michael E. Rule¹, Timothy O’Leary¹, and Christopher D. Harvey²

¹Department of Engineering, University of Cambridge, Cambridge CB21PZ, United Kingdom

²Department of Neurobiology, Harvard Medical School, Boston, MA 02115 USA

August 13, 2019

Abstract

The nervous system learns new associations while maintaining memories over long periods, exhibiting a balance between flexibility and stability. Recent experiments reveal that neuronal representations of learned sensorimotor tasks continually change over days and weeks, even after animals have achieved expert behavioral performance. How is learned information stored to allow consistent behavior despite ongoing changes in neuronal activity? What functions could ongoing reconfiguration serve? We highlight recent experimental evidence for such representational drift in sensorimotor systems, and discuss how this fits into a framework of distributed population codes. We identify recent theoretical work that suggests computational roles for drift and argue that the recurrent and distributed nature of sensorimotor representations permits drift while limiting disruptive effects. We propose that representational drift may create error signals between interconnected brain regions that can be used to keep neural codes consistent in the presence of continual change. These concepts suggest experimental and theoretical approaches to studying both learning and maintenance of distributed and adaptive population codes.

Highlights

- Experimental advances allow us to see long-term drift in neural representations
- Drift challenges classical notions of receptive fields and engrams
- Drift is consistent with distributed population codes
- Distributed error signals across brain areas could help maintain such codes

Introduction

Heraclitus of Ephesus is quoted as saying that one cannot step into the same river twice¹. Accordingly, our brains continually renew their molecular and cellular components, and the neuronal substrates of our experiences and memories are subject to continual turnover [1,2,3]. Such turnover could occur without changing the relationship between neuronal activation and the external world. However, recent experiments reveal continual reorganization of neuronal responses in circuits essential for specific tasks, even when tasks are fully learned [4,5,6,7,8].

This apparent instability challenges the view that synaptic connectivity and individual neuronal responses correlate directly with memory. Can we reconcile stable behavior with apparent instability

¹Plato, Cratylus, 360 B.C.E

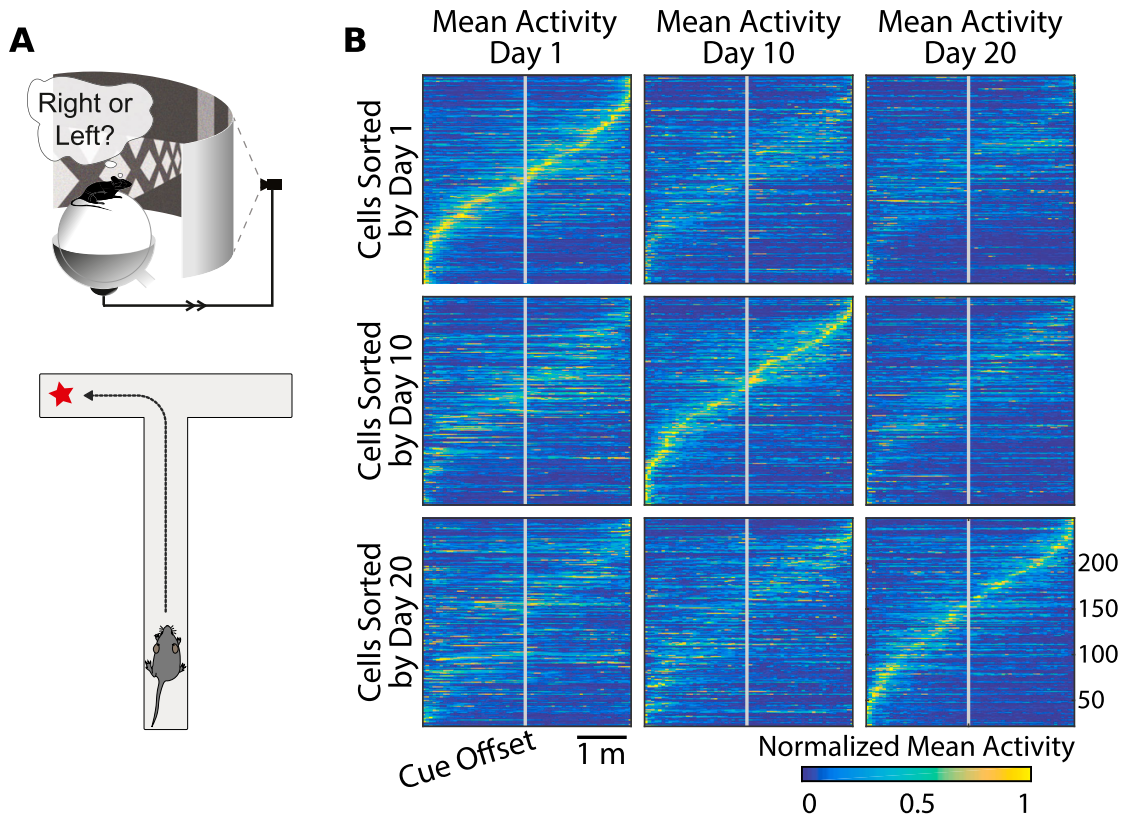


Figure 1: **Coding of spatial navigation in Posterior Parietal Cortex (PPC) drifts over days** (adapted from Driscoll et al. [4]). (A) Driscoll et al. [4] placed mice in a virtual reality environment, and required that subject remember visual cues to navigate to a target. Population activity was recorded with single-neuron resolution over days using calcium fluorescence imaging. (B) Raster plots showing average calcium signals from several hundred PPC neurons imaged over multiple days, with task location on the horizontal axis. Each row corresponds to a neuron, and mean activity is represented by color. Location-dependent activation drifted slowly over days: single neurons gained and lost location sensitivity or changed their tuning. Sorting cells by activation on any given day reveals population coding of maze location.

in behavior-related neuronal activity? Experimental examples of stability and instability in neuronal representations have been extensively reviewed previously [9,10,11]. In this review, we focus on recent and established theoretical models that address this problem, including potential functional roles of continual circuit reconfiguration. We suggest experimental and theoretical strategies to study how and why brain circuits continually evolve during stable behavior.

Experiments find consistent population patterns in the presence of single-neuron drift

Recent experiments have found that neuronal representations of familiar environments and learned tasks reconfigure or ‘drift’ over time [4,6,7,8]. Here we take ‘representations’ to mean neural activity that is correlated with task-related stimuli, actions, and cognitive variables. Representations could include, for example, single-cell receptive fields in sensory areas, or population activity vectors that guide behavior. We use the term ‘representational drift’ to describe ongoing changes in these representations that occur without obvious changes in behavior.

We will highlight one recent example to illustrate representational drift. Driscoll and colleagues [4] designed a sensorimotor task in a virtual reality environment, in which a mouse was trained to navigate a T-maze (Figure 1a). For each trial, the mouse was presented with a visual cue, which instructed it whether to turn left or right at the end of the maze to receive a reward. Mice performed this task at greater than 90% accuracy for weeks. Using chronic two-photon calcium imaging, the authors monitored the activity of large groups of individual neurons in the posterior parietal cortex (PPC), which is

known to be required for solving the task [4,12]. Neurons tended to be transiently active during task trials, with different neurons active at different parts of the trial. This forms a sequence of neuronal activity across the population that tiles the task (Figure 1b, diagonal panels). We refer to this activity sequence as a representation of the task.

Crucially, Driscoll and colleagues found that the PPC representation was not stable over multiple days and weeks. As shown in each row of Figure 1b, the same neurons exhibited markedly different activation patterns on different days. The most common change was that neurons had altered levels of activity and thus exited or entered the population representation. Less frequently, cells exhibited changes in selectivity. Over weeks, the task-related activity in PPC had nearly entirely reconfigured, but on any given day a subset of the population could be identified that tiled the task (Fig. 1b, diagonals). Each animal's task performance remained consistently high and behavior was not measurably altered by representational drift.

Similar types of drift have been reported in a number of brain areas, including the hippocampus and sensory and motor parts of neocortex [6,7,8,13]. In addition, there is widespread evidence for surprising degrees of structural plasticity in dendritic spines [7,8,2]. For example, in the hippocampus, all dendritic spines are expected to turn over in the period of several weeks [7]. Such dramatic synapse turnover suggests that circuits are continually rewiring even though animals can maintain stable task performance and memories. We emphasize that drift is not observed in all brain areas and for all tasks [14,15]. Nevertheless, the finding of representational drift raises profound questions about how behavior is learned and controlled in neural circuits, and what constitutes a memory of such learned behavior.

Distributed population codes can accommodate representational drift

Representational drift might appear problematic for long-term encoding of memories and associations. However, redundant representations may allow some level of drift without disrupting behavior. Even in simple nervous systems, the existence of circuit configurations with different anatomical connectivity or physiological profiles but similar overall function is well documented [16]. Redundancy is often considered to be a biological necessity because brains must be robust to failure in individual neurons and to environmental perturbations. The brain may therefore achieve robustness via degeneracy, where high-dimensional representations preserve behavior while allowing for a vast number of equivalent circuit configurations to be realized [17].

There is considerable evidence that the brain employs high-dimensional representations of inherently low dimensional tasks [18,19,20,21]. A low dimensional task can be represented in higher dimensional population activity in a variety of configurations. To illustrate, we can explore the neuronal population representation of the task from Driscoll et al. by applying dimensionality reduction to PPC population activity. In this example an unsupervised dimensionality-reduction algorithm [22,23] is used to find 2D projections of population activity that preserve nearest-neighbor structure in population activity. Without knowing the details of the task or observed location, this algorithm identifies a 'T-shaped' cloud of population activity states (Figure 2a). Each point in the cloud corresponds to the population activity at a single time bin in the trial, and collectively the cloud of points maps out the animal's navigational trajectories during the task. Although internally consistent neuronal representations can be identified (Figure 2b), the way that single neurons encode such representations changes over time.

The low-dimensional structure extracted in Figure 2a sits in a much higher dimensional space of population activity. There are potentially many degrees of freedom for this structure to move around in population activity space while retaining the topology and local structure of the T shape. Such movement could accommodate different contributions from different neurons across time, or changes in single-cell tuning.

Interestingly, such high dimensional representations have other, less obvious benefits. A recent study by Raman and colleagues [24] showed that networks with excess connectivity can learn more

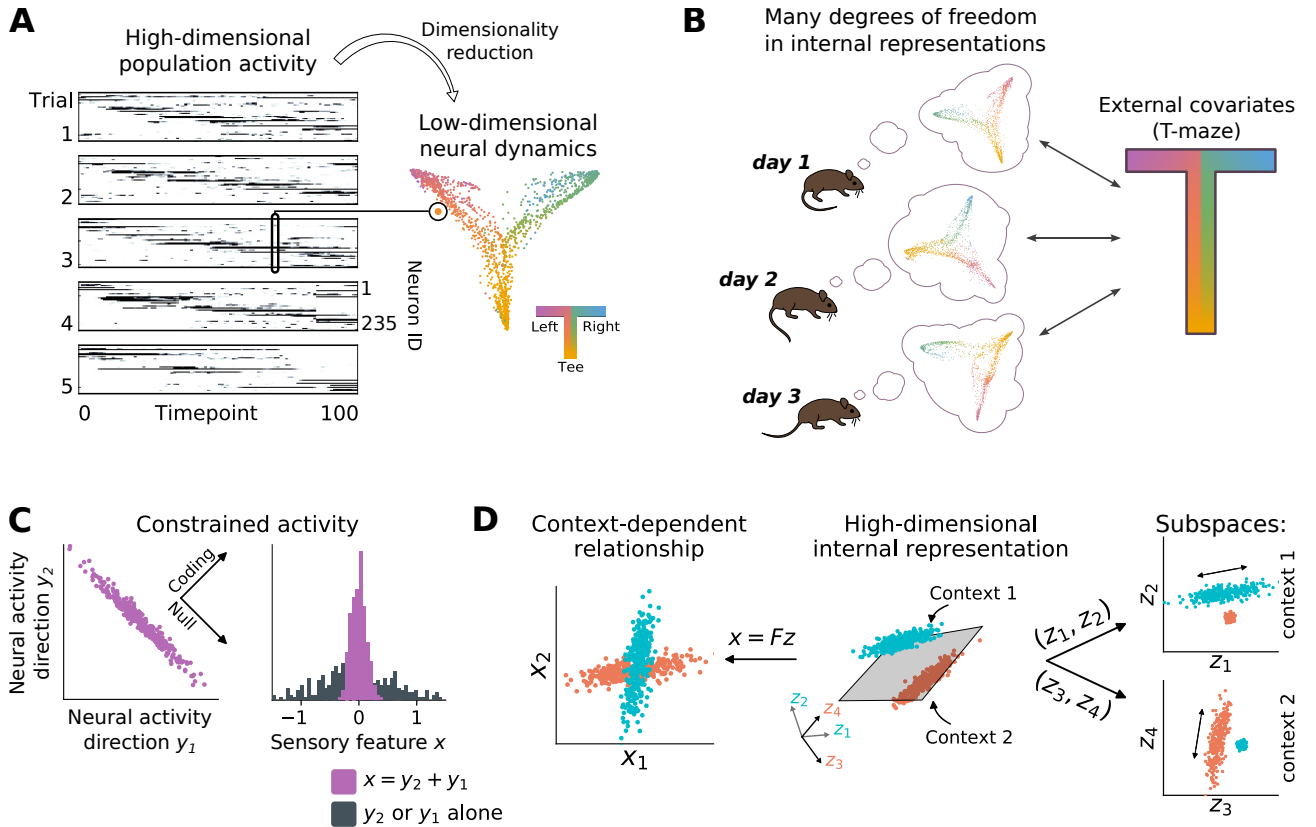


Figure 2: Internal representations have unconstrained degrees of freedom that allow drift. (A) Nonlinear dimensionality reduction of population activity recovers the low-dimensional structure of the T-maze in [4] (Figure 1a). Each point represents a single time-point of population activity, and is colored according to location in the maze. (B) Point clouds illustrate low-dimensional projections of neural activity as in (a). Although unsupervised dimensionality-reduction methods can recover the task structure on each day, the way in which this structure is encoded in the population can change over days to weeks. (C) *Left:* Neural populations can encode information in relative firing rates and correlations, illustrated here as a sensory variable encoded in the sum of two neural signals ($y_1 + y_2$). Points represent neural activity during a repeated presentation of the same stimulus. Variability orthogonal to this coding axis does not disrupt coding, but could appear as drift in experiments if it occurred on slow timescales. *Right:* Such distributed codes may be hard to read-out from recorded sub-populations (e.g. y_1 or y_2 alone; black), especially if they entail correlations between brain areas. (D) *Left:* External covariates may exhibit context-dependent relationships. Each point here reflects a neural population state at a given time-point. The relationship between directions x_1 and x_2 changes depending on context (cyan vs. red). *Middle:* Internally, this can be represented a mixture model, in which different subspaces are allocated to encode each context, and the representations are linearly-separable (gray plane). *Right:* The expanded representation contains two orthogonal subspaces that each encode a separate, context-dependent relationship. This dimensionality expansions increases the degrees of freedom in internal representations, thereby increasing opportunities for drift.

rapidly and to a higher asymptotic performance on tasks of fixed complexity. A high-dimensional representation can therefore be advantageous for learning.

A notable idea that emerges from the high-dimensional encoding of low-dimensional tasks is that of a ‘null space’. The null space is a subspace of population activity that is orthogonal to a low-dimensional task representation [25,26]. This is illustrated in Figure 2c, which depicts how population activity in two sub-populations, y_1 and y_2 , might co-vary in a population that encodes a specific feature, x . If we suppose that the feature is encoded in the sum of the activity, $y_1 + y_2$, then tight tuning with respect to x can coexist with a large variation in a null direction (the $y_1 - y_2$ direction). Sub-populations may appear variable, even if the global representation is well constrained. For example, recent work has suggested that the population code in V1 is structured, with distributed representations having lower variance than individual neurons [27].

The null-space concept has been used to explain how a single neuronal population can represent

multiple behavioral contexts, for example motor planning and execution [28,26,29]. To illustrate this, Figure 2d depicts samples of activity from a single population during two different behavioral contexts. Behavioral variables, x_1 and x_2 , have different relationships to each other in each context (left panel). A high-dimensional representation in neuronal activity, z_i allows these contexts to occupy different parts of population activity space (middle panel). By examining subspaces of the population activity, one could observe correlates of task variables in one context, but not the other (right panel). This implies that neuronal activity could drift in directions unrelated to encoding or task performance [25,30]. If the dimensionality of population activity is much higher than the dimensionality of the task, even random drift will lie mostly within this ‘null space’. Thus, high-dimensional population representations can tolerate drift and allow multiple circuit configurations to lead to similar outputs.

Further evidence of distributed and redundant population representations has emerged from recent work highlighting the distributed nature of sensorimotor information [31]. In particular, recent studies have shown that motor outputs influence sensory encoding [32,33]. Moreover, recent reports show that stimulus, action, and cognitive variables are distributed throughout sensory and motor areas, often in overlapping representations. Reports of this type are too numerous to list here (but see for example [31,34,35,36]).

The presence of widely-distributed representations thus necessitates understanding drift at a wider neuronal population level, even across brain areas, than is typically examined in experiments. This wider examination will be essential to understand the scale of drift in the representations relative to the global population representation, including to correctly identify coding dimensions of distributed activity.

Representational drift may be inevitable in distributed, adaptive circuits

Representational drift is sometimes considered as a passive, noise-driven process. However, it could also reflect other ongoing processes such as learning. In typical lab experiments, a specific task of interest and its associated representations are studied, but the same population of neurons is likely used for other aspects of the subject’s life. Thus, over the course of an experiment, animals likely learn many new associations and have new experiences, which must be incorporated into the neuronal populations being studied [37,38]. To prevent new associations from disrupting previously learned associations, the brain may need to re-encode them.

In the work of Driscoll et al., these ideas were explored by training mice to learn new sensorimotor associations after they had already learned earlier associations. Interestingly, they found that the same neurons appeared to be used for the representations of previously learned associations and for the development of new associations during learning. This finding demonstrates that representational drift could indeed reflect new learning and suggests that a neuronal population can simultaneously be utilized for learning and memory. This idea of drift as ongoing learning is consistent with recent theoretical work that predicts a highly plastic subset of neurons attuned to population activity [39].

Even in the absence of explicit learning, neuronal representations continually adapt to encode information efficiently [40]. As sensory representations adjust, downstream areas must also adjust either their connectivity or internal representations to remain consistent. Efficient coding is not limited to sensory functions: even sensory areas learn to anticipate motor output [41], and one might expect networks to track shifts in environmental statistics, including evolving cognitive and memory effects. These factors contribute to the neural ‘environment’ being in perpetual flux. Drift may therefore be an expected consequence of ongoing refinement and consolidation.

Predictive coding and internal error signals could detect and correct drift

Regardless of the sources of drift, neuronal representations must remain within the subspace of equivalent representations in order to preserve behaviors and learned associations (depicted in Figure 3a as a surface). Over time, changes in neuronal representations are expected to accumulate, eventually leading to disruptive effects that cannot be accommodated by redundancy. Therefore, some continual corrective action is needed constrain neuronal representations on long timescales. In many situations, external stimuli could provide this corrective feedback. For example, in the work from Driscoll et al., the mouse was rewarded after each correct trial, which could serve as an external signal to update the subspace of adequate neuronal representations for behavior. In addition, other mechanisms for maintaining coherent representations could also be used, including off-line rehearsal [42] and reactivation of cell assemblies [43].

In the absence of external learning cues, we propose that internal error signals exchanged between recurrently-connected brain regions could maintain consistency in distributed representations. For example, spatial navigation requires consistent representations throughout sensory, association, and motor regions. A change in representation in any one area could disrupt consistency in representations with other regions. Plasticity in the other regions could then be used to compensate for this change such that representations distributed across brain regions would drift in a consistent manner (Figure

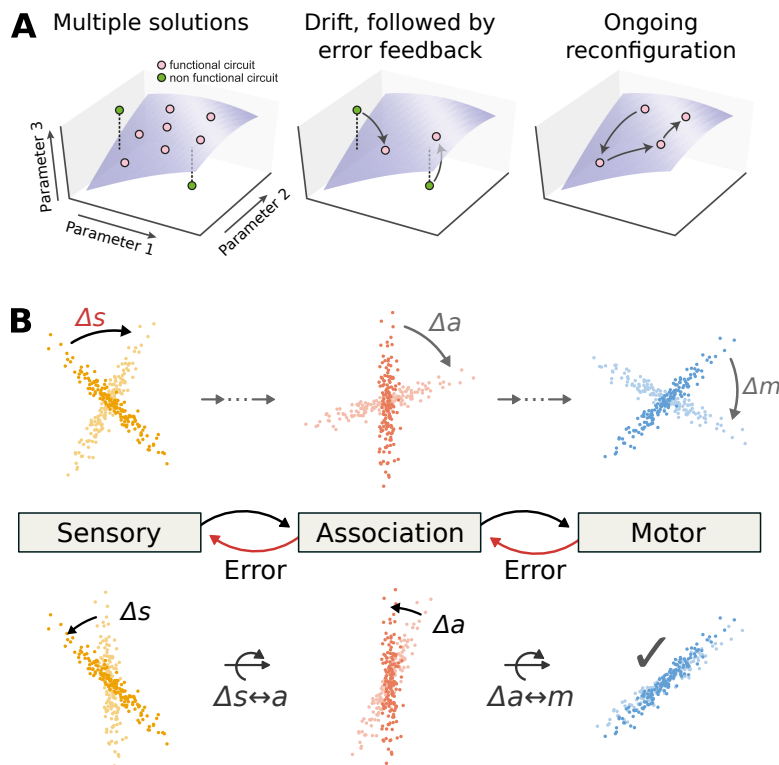


Figure 3: Local changes in recurrent networks have global effects, and global processes can compensate. (A) The curved surfaces represent network configurations suitable for a given sensorimotor task, i.e. neural connections and tunings that generate a consistent behavior. Each axis represents different circuit parameters. Ongoing processes that disrupt performance must be corrected via error feedback to maintain overall sensorimotor accuracy (B) Colored dots represent projections of neural population activity onto task-relevant dimensions at various time-points. Activity is illustrated in three hypothetical areas, depicting a feed-forward transformation of a stimulus input into a motor output. (*top*) If the representation in one area changes (e.g. rotation of an internal sensory representation, Δs , curved black arrow), downstream areas must also compensate to avoid errors (e.g. motor errors, Δm , curved gray arrows). (*bottom*) Although the original perturbation was localized, compensation can be distributed over many areas. Each downstream area can adjust how it interprets its input. This is illustrated here as curved arrows, which denote a compensatory rotation that partially corrects the original perturbation. The distributed adjustment in neural tuning may appear as drift to experiments that examine only a local sub-population.

3b).

This concept fits within the framework of predictive coding, in which neural circuits learn to predict the activity patterns in one another with the goal of minimizing internal prediction errors [44,45]. In this framework, one brain region might generate an error signal if the input it receives from another region is different than expected, such as due to drift in the input area. Such an error signal could guide plasticity to maintain consistency between representations across areas. Much work has highlighted the concept of predictive coding and error signals in the context of comparisons between internal predictions and incoming sensory signals [33,46]. However, to our knowledge, this concept has not been explored in the context of extensive internal predictions between brain regions. This will be an interesting area for future research, in particular to identify if such error signals exist and to test how these signals could be used to maintain coherent representations across areas.

Possible computational uses of representational drift

Drift and instability in neuronal representations could serve important computational functions. Some insights into these potential functions come from comparisons to strategies used in machine learning. For example, a common experimental finding is that some neurons go from being highly active in a task- or behavior-relevant manner to being silent on subsequent days. The transient silencing of single neurons (e.g. [47]) could be a neural correlate of the “drop-out” training strategy used to regularize deep neural networks [48].

Recent work proposed that drift allows for ‘time-stamping’ of events, following the observation that different sets of hippocampal place cells are active in an environment on different days [5]. The set of active place cells conveys information about not only the environment but also time. Time-stamping could support episodic memory, disambiguating similar environmental contexts separated in time. Accordingly, mutually-exclusive population representations of distinct memories is also observed on fast timescales [49], and temporal context appears to be involved in episodic memory at fast timescales [50]. This connects, at an abstract level, to the recently proposed machine-learning strategy ‘context-dependent gating’ [51], which silences subsets of neurons in a context-dependent manner in order to attenuate interference. If time itself is an important contextual variable, then distinct contextual representations could emerge naturally from drifting representations.

Recent theoretical work has suggested that drift may allow the brain to sample from a large space of possible solutions [52]. In this case, learning and drift work together to move toward optimal solutions while sampling enough possibilities to avoid globally-suboptimal local solutions. In this case, drift could be a deliberate strategy to sample the configuration landscape or a byproduct of noisy and error-prone learning. Some theories indicate that fluctuations are an expected feature of optimization in noisy systems [25] and that drift may therefore support stochastic reinforcement learning Kappel et al. [53].

Outlook

The brain is an adaptive system, and its structure therefore changes. While this has been appreciated in the context of learning, recent experimental findings suggest something stronger: some parts of the brain cannot remain fixed, even in experimental paradigms designed to study stable behavior. A neuron that is several synaptic connections away from a sensory input or motor output is only weakly tethered to the external world. Neurons that participate in abstract representations and high-level behavioral plans are therefore free to reconfigure within limits set by the degeneracy of the neural code at the circuit level.

This realization suggests approaches for capturing invariant structures in population activity that underlie stable sensorimotor behavior. It also implies that internal feedback signals between brain areas are pervasive. This provides a framework for theoretical models of neural circuits and may help understand the logic of connectivity in many brain areas. Integrating theories of collective neural dynamics,

learning, and predictive coding in distributed representations is therefore essential to understand how sensorimotor representations evolve.

We propose experimental and theoretical shifts in how we consider learning and memory. Rather than viewing learning and memory as sequential, and potentially discrete, events in separate circuits, we propose that it is important to study them together to understand their interaction. The brain is an interconnected network, and changes in one area likely influence distant neuronal populations. Globally coordinated plasticity may be needed to preserve existing associations. In other words, “to stay the same, everything must change”². Experiments that track the interactions between brain regions will therefore be essential to examine long-term neuronal population dynamics during learning and memory as well as during stable behavior. Such experiments will test emerging theories of population codes and memory in the presence of constant change, revealing how the brain achieves one of its most essential functions—reorganizing with experience to guide future actions.

1 Acknowledgements

We thank Dhruva Raman, Sofia Soares, Thomas Burger and Monika Jozsa for comments on the manuscript. This work is supported by the Human Frontier Science Program, ERC grant StG 716643 FLEXNEURO, and NIH grants (NS108410, NS089521, MH107620).

2 Special and outstanding interest papers

- ****Driscoll et al. [4]** The authors examine neural representations for spatial navigation in mouse posterior parietal cortex using a closed-loop virtual reality environment, and find that the neural code drifts and reconfigures itself over days.
- ****Attinger et al. [33]** Apical dendrites in layer 2/3 of visual cortex receive rapid feedback signals related to sensorimotor errors. This work argues for an experience-dependent internal model, and the possibility of global and distributed error feedback when that model is violated.
- ****Montijn et al. [27]** The authors show that distributed population relationships in visual cortex encode stimuli more robustly than single-neurons or pairwise statistics, demonstrating that apparent variability may be due to subsampling population dynamics.
- ****Gallego et al. [21]** This study demonstrates that sensorimotor processing is best viewed in terms of models of population activity lying on low-dimensional manifolds, and that population-level correlations reflect stable invariants underlying motor control across diverse tasks.
- ****Kappel et al. [53]** This computational study shows that instability in single-neurons and synapses is compatible with sustained sensorimotor performance, provided continued error feedback is available. Once the task is learned, noise and other biological processes cause the network configuration to drift randomly, rearranging as it explores the space of equivalent solutions.
- ***Katlowitz et al. [14]** This study finds extremely robust and long-lived neural representations underlying birdsong, in contrast to many studies that report single-neuron instability.
- ***Prsa et al. [47]** While studying learning and conditioning in single neurons, this study finds that neurons in mouse motor cortex drop out of the population response on the scale of days.
- ***Masse et al. [51]** This computational study finds that silencing parts of a neural network in a context-dependent manner can allow it to learn multiple tasks with reduced interference.
- ***Chalk et al. [44]** This theoretical work explores how sensory systems might efficiently encode the future, and predicts surprising population responses in a model of visual encoding.

²“Se vogliamo che tutto rimanga com’è, bisogna che tutto cambi”, Giuseppe Tomasi di Lampedusa, *Il Gattopardo*, 1958

3 Bibliography

References

- [1] Susumu Tonegawa, Michele Pignatelli, Dheeraj S Roy, and Tomas J Ryan. Memory engram storage and retrieval. *Current opinion in neurobiology*, 35:101–109, 2015.
- [2] Gianluigi Mongillo, Simon Rumpel, and Yonatan Loewenstein. Intrinsic volatility of synaptic connections—a challenge to the synaptic trace theory of memory. *Current opinion in neurobiology*, 46:7–13, 2017.
- [3] Simon Rumpel and Jochen Triesch. The dynamic connectome. *e-Neuroforum*, 22(3):48–53, 2016.
- [4] Laura N Driscoll, Noah L Pettit, Matthias Minderer, Selmaan N Chettih, and Christopher D Harvey. Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, 170(5):986–999, 2017.
- [5] Alon Rubin, Nitzan Geva, Liron Sheintuch, and Yaniv Ziv. Hippocampal ensemble dynamics times-tamp events in long-term memory. *eLife*, 4:e12247, 2015.
- [6] Yaniv Ziv, Laurie D Burns, Eric D Cocker, Elizabeth O Hamel, Kunal K Ghosh, Lacey J Kitch, Abbas El Gamal, and Mark J Schnitzer. Long-term dynamics of CA1 hippocampal place codes. *Nature neuroscience*, 16(3):264, 2013.
- [7] Alessio Attardo, James E Fitzgerald, and Mark J Schnitzer. Impermanence of dendritic spines in live adult CA1 hippocampus. *Nature*, 523(7562):592, 2015.
- [8] Alessio Attardo, Ju Lu, Takashi Kawashima, Hiroyuki Okuno, James E Fitzgerald, Haruhiko Bito, and Mark J Schnitzer. Long-term consolidation of ensemble neural plasticity patterns in hippocampal area CA1. *Cell reports*, 25(3):640–650, 2018.
- [9] Claudia Clopath, Tobias Bonhoeffer, Mark Hübener, and Tobias Rose. Variance and invariance of neuronal long-term representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):20160161, 2017.
- [10] Henry Lütcke, David J Margolis, and Fritjof Helmchen. Steady or changing? long-term monitoring of neuronal population activity. *Trends in neurosciences*, 36(7):375–384, 2013.
- [11] Milton Llera-Montero, João Sacramento, and Rui Ponte Costa. Computational roles of plastic probabilistic synapses. *Current Opinion in Neurobiology*, 54:90 – 97, 2019. ISSN 0959-4388. Neurobiology of Learning and Plasticity.
- [12] Christopher D Harvey, Philip Coen, and David W Tank. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62, 2012.
- [13] Daniel Huber, Diego A Gutnisky, Simon Peron, Daniel H O’connor, J Simon Wiegert, Lin Tian, Thomas G Oertner, Loren L Looger, and Karel Svoboda. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature*, 484(7395):473, 2012.
- [14] Kalman A Katlowitz, Michel A Picardo, and Michael A Long. Stable sequential activity underlying the maintenance of a precisely executed skilled behavior. *Neuron*, 2018.
- [15] Ashesh K Dhawale, Rajesh Poddar, Steffen BE Wolff, Valentin A Normand, Evi Kopelowitz, and Bence P Ölveczky. Automated long-term recording and analysis of neural activity in behaving animals. *eLife*, 6:e27702, 2017.
- [16] Eve Marder, Marie L Goeritz, and Adriane G Otopalik. Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Current opinion in neurobiology*, 31:156–163, 2015.

- [17] Braden AW Brinkman, Fred Rieke, Eric Shea-Brown, and Michael A Buice. Predicting how and when hidden neurons skew measured synaptic interactions. *PLoS computational biology*, 14(10): e1006490, 2018.
- [18] John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500, 2014.
- [19] Carlos E Vargas-Irwin, Lachlan Franquemont, Michael J Black, and John P Donoghue. Linking objects to actions: encoding of target object and grasping strategy in primate ventral premotor cortex. *Journal of Neuroscience*, 35(30):10888–10897, 2015.
- [20] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. *Neuron*, 94(5):978–984, 2017.
- [21] Juan A Gallego, Matthew G Perich, Stephanie N Naufel, Christian Ethier, Sara A Solla, and Lee E Miller. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature communications*, 9(1):4233, 2018.
- [22] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
- [23] Alon Rubin, Liron Sheintuch, Noa Brande-Eilat, Or Pinchasof, Yoav Rechavi, Nitzan Geva, and Yaniv Ziv. Revealing neural correlates of behavior without behavioral measurements. *bioRxiv*, 2019. doi: 10.1101/540195.
- [24] Dhruva V Raman, Adriana P Rotondo, and Timothy O’Leary. Fundamental bounds on learning performance in neural circuits. *Proceedings of the National Academy of Sciences*, 116(21):10537–10546, 2019.
- [25] Uri Rokni, Andrew G Richardson, Emilio Bizzi, and H Sebastian Seung. Motor learning with unstable neural representations. *Neuron*, 54(4):653–666, 2007.
- [26] Matthew T Kaufman, Mark M Churchland, Stephen I Ryu, and Krishna V Shenoy. Cortical activity in the null space: permitting preparation without movement. *Nature neuroscience*, 17(3):440, 2014.
- [27] Jorrit S Montijn, Guido T Meijer, Carien S Lansink, and Cyriel MA Pennartz. Population-level neural codes are robust to single-neuron variability from a multidimensional coding perspective. *Cell reports*, 16(9):2486–2498, 2016.
- [28] Gamaleldin F Elsayed, Antonio H Lara, Matthew T Kaufman, Mark M Churchland, and John P Cunningham. Reorganization between preparatory and movement population responses in motor cortex. *Nature communications*, 7:13239, 2016.
- [29] Carlos E Vargas-Irwin, Jessica M Feldman, Brandon King, John D Simeral, Brittany L Sorice, Erin M Oakley, Sydney S Cash, Emad N Eskandar, Gerhard M Friehs, Leigh R Hochberg, et al. Watch, imagine, attempt: Motor cortex single unit activity reveals context-dependent movement encoding in humans with tetraplegia. *Frontiers in human neuroscience*, 12:450, 2018.
- [30] Shaul Druckmann and Dmitri B Chklovskii. Neuronal circuits underlying persistent representations despite time varying activity. *Current Biology*, 22(22):2095–2103, 2012.
- [31] Matthias Minderer, Kristen D Brown, and Christopher D Harvey. The spatial structure of neural encoding in mouse posterior cortex during navigation. *Neuron*, 102(1):232–248, 2019.
- [32] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brain-wide population activity. *bioRxiv*, page 306019, 2018.

- [33] Alexander Attinger, Bo Wang, and Georg B Keller. Visuomotor coupling shapes the functional development of mouse visual cortex. *Cell*, 169(7):1291–1302, 2017.
- [34] Dmitriy Aronov, Rhino Nevers, and David W Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719, 2017.
- [35] Janelle MP Pakan, Valerio Francioni, and Nathalie L Rochefort. Action and learning shape the activity of neuronal circuits in the visual cortex. *Current opinion in neurobiology*, 52:88–97, 2018.
- [36] Aman B Saleem, E Mika Diamanti, Julien Fournier, Kenneth D Harris, and Matteo Carandini. Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*, 562(7725):124, 2018.
- [37] Yadin Dudai, Avi Karni, and Jan Born. The consolidation and transformation of memory. *Neuron*, 88(1):20–32, 2015.
- [38] Arun Asok, Félix Leroy, Joseph B Rayman, and Eric R Kandel. Molecular mechanisms of the memory trace. *Trends in neurosciences*, 2018.
- [39] Yann Sweeney and Claudia Clopath. Population coupling predicts the plasticity of stimulus responses in cortical circuits. *bioRxiv*, page 265041, 2018.
- [40] Meytal Wilf, Francesca Strappini, Tal Golan, Avital Hahamy, Michal Harel, and Rafael Malach. Spontaneously emerging patterns in human visual cortex reflect responses to naturalistic sensory stimuli. *Cerebral cortex*, 27(1):750–763, 2017.
- [41] Rohan N Ramesh, Christian R Burgess, Arthur U Sugden, Michael Gyetvan, and Mark L Andermann. Intermingled ensembles in visual association cortex encode stimulus identity or predicted outcome. *Neuron*, 100(4):900–915, 2018.
- [42] Szabolcs Káli and Peter Dayan. Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature neuroscience*, 7(3):286, 2004.
- [43] Michael Jan Fauth and Mark CW van Rossum. Self-organized reactivation maintains and reinforces memories despite synaptic turnover. *eLife*, 8:e43717, 2019.
- [44] Matthew Chalk, Olivier Marre, and Gašper Tkačič. Toward a unified theory of efficient, predictive, and sparse coding. *Proceedings of the National Academy of Sciences*, 115(1):186–191, 2018.
- [45] Veronika Koren and Sophie Denève. Computational account of spontaneous activity as a signature of predictive coding. *PLoS computational biology*, 13(1):e1005355, 2017.
- [46] Georg B Keller and Thomas D Mrsic-Flogel. Predictive processing: a canonical cortical computation. *Neuron*, 100(2):424–435, 2018.
- [47] Mario Prsa, Gregorio L Galiñanes, and Daniel Huber. Rapid integration of artificial sensory feedback during operant conditioning of motor cortex neurons. *Neuron*, 93(4):929–939, 2017.
- [48] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [49] Arnaud Malvache, Susanne Reichinnek, Vincent Villette, Caroline Haimerl, and Rosa Cossart. Awake hippocampal reactivations project onto orthogonal neuronal assemblies. *Science*, 353(6305):1280–1283, 2016.
- [50] Nicole M Long and Michael J Kahana. Hippocampal contributions to serial-order memory. *Hippocampus*, 29(3):252–259, 2019.

- [51] Nicolas Y. Masse, Gregory D. Grant, and David J. Freedman. Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proceedings of the National Academy of Sciences*, 115(44):E10467–E10475, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1803839115.
- [52] David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Network plasticity as bayesian inference. *PLoS computational biology*, 11(11):e1004485, 2015.
- [53] David Kappel, Robert Legenstein, Stefan Habenschuss, Michael Hsieh, and Wolfgang Maass. A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning. *eNeuro*, 5(2):ENEURO–0301, 2018.