Cognition and the single neuron: How cell types construct the dynamic computations of frontal cortex

Amelia J. Christensen[1], **Torben Ott**[1,2], **Adam Kepecs**[1]

[1]Department of Neuroscience and Department of Psychiatry, Washington University in St. Louis, St. Louis, MO 63110, USA

[2]Bernstein Center for Computational Neuroscience Berlin, Humboldt University of Berlin, Berlin, Germany

Abstract

Frontal cortex is thought to underlie many advanced cognitive capacities, from self-control to long term planning. Reflecting these diverse demands, frontal neural activity is notoriously idiosyncratic, with tuning properties that are correlated with endless numbers of behavioral and task features. This menagerie of tuning has made it difficult to extract organizing principles that govern frontal neural activity. Here, we contrast two successful yet seemingly incompatible approaches that have begun to address this challenge. Inspired by the indecipherability of single-neuron tuning, the first approach casts frontal computations as dynamical trajectories traversed by arbitrary mixtures of neurons. The second approach, by contrast, attempts to explain the functional diversity of frontal activity with the biological diversity of cortical cell-types. Motivated by the recent discovery of functional clusters in frontal neurons, we propose a consilience between these population and cell-type-specific approaches to neural computations, advancing the conjecture that evolutionarily inherited cell-type constraints create the scaffold within which frontal population dynamics must operate.

Introduction

Frontal cortex is thought to underlie a wide variety of cognitive capacities–ethologically crucial functions like long-term planning [1], short-term memory, cognitive control [2], social cognition [3], and self-reflection [4]. More generally, frontal cortex acts to monitor, combine, modulate, and control activity across brain regions, on a variety of timescales, and its involvement in high-level behavior has been summarized as an additional layer of hierarchical processing within neocortex. Indeed, a wealth of studies have found that the activity of frontal cortical neurons correlates with a seemingly endless number of features–from sensory and spatial information to decisions and economic value to abstract concepts and rules. Is it possible to organize these diverse neural activity patterns into a core set of

Corresponding author: Kepecs, Adam (akepecs@wustl.edu) ameliac@wustl.edu (A.J. Christensen), torbenott@wustl.edu (T. Ott).

Conflict of interest
None.

specific roles frontal cortical neurons perform in support of the computations involved in higher cognitive behaviors?

In visual cortex, a neuron's tuning curve, that is, the correlation of neural activity with external features, is suggestive of the operation that neuron might perform. A visual cortical neuron responds only to specific visual features (like edges) in a certain region of retinotopic space, and its computational role can be described as filtering the incoming sensory data through its receptive fields [5]. Although this description is incomplete, it has inspired convolutional neural networks [6], models that not only predict the activity of visual cortical neurons [7], but have also revolutionized computer vision. Alas, the neural tuning approach has not led to the discovery of the analogous neural roles in higher cognition, or a new deep learning revolution. Rather it has just led to a taxonomy of occupants for the cortical-tuning zoo. So then, how can we begin to search for organizing principles that reveal the functions of these frontal neurons?

This question is now being tackled by new approaches, with major breakthroughs coming from two disparate directions. One approach focuses on organizing frontal neural dynamics by looking for patterns in population neural activity [8–10]. Separately, another approach focuses on classifying single-cortical-neurons into different anatomically and genetically defined cell types [11]. After reviewing recent insights from both approaches, we argue that cell-type specific connectivity motifs constrain the computational role played by individual frontal neurons.

Inferring computations from neural population dynamics

Classic approaches for interpreting neural activity rely on analysis of a subset of recorded single neuron firing rates by aligning the activity to relevant task phases and condition-averaging, to create peri-stimulus time histograms (PSTHs). While this approach has had famous success summarizing neural activity relatively close to the sensory periphery, frontal PSTHs can display baffling complexity [12]. That is, coding for task-relevant variables are randomly mixed throughout the neural population, and only by considering the contributions of many neurons together can the state of the external variable(s) be consistently read out from the neural activity [13]. This "mixed selectivity" has been observed in a range of frontal [12,14] and parietal [15] areas. Thus, facilitated by the advent of new technologies to record, and data analysis techniques to analyze large populations of neurons, considering neural populations as a unit has become the dominant approach of systems neuroscience across the brain, and particularly in frontal cortex.

As a brief background, population dynamics can be represented as trajectories in a high-dimensional space, with axes defined by the firing rate of each neuron [12]. By themselves, these high dimensional trajectories are simply a change of basis and provide no additional information over a list of classical PSTHs. However, combined with techniques to find lower-dimensional subspaces or manifolds in which neural trajectories move (collectively referred to as dimensionality reduction [16]), considering neural populations as a trajectory has inspired the application of mathematical techniques (like dynamical systems [8], geometry [17], and topology [18]) to search for patterns in neural data (Figure 1a). Population approaches have many advantages over single-neuron analyses–they often have

less selection bias (because all or most recorded neurons are considered), they are applicable online and in single trials, and they have enabled brain-computer-interface based motor-control and communication for severely disabled human patients [19]. The population trajectory approach was popularized by seminal contributions in the field of motor learning [20], preparation [21], and control [22]; summarized elsewhere [9,13,23].

The neural population approach has revealed dynamical operations that are generally useful for flexible decision-making. Flexible, context-dependent sensory decision making is a prototypical case where mixed selectivity dominates single-neuron responses [12]. In context-dependent decision making tasks, as opposed to being represented by single neurons, context has been observed to change the shape of the neural activity landscape, to preferentially bias application of one decision rule over another [14,24–27] (Figure 1b). That is, if we imagine frontal cortex as a funnel into which various sensory variables are poured, and out of which comes a decision, context can change the shape of the funnel, effectively biasing which decision the trajectories roll towards. This context dependent-change in dynamical landscape is an emerging motif often found in frontal cortical operations. Rather than changing the shape of the neural funnel, computational flexibility can also be achieved by changing the inputs or initial conditions (e.g., where the sensory-balls start out inside the funnel) of the system dynamics [28] (Figure 1). This computational strategy has been observed in interval timing reproduction tasks [29–32], where afferent and intrinsic input play the role of setting initial conditions of population dynamics.

Although we focused on studies where computation was performed through alteration of the input or shape of the neural state space, the population dynamics approach, combined with the simplifying assumption of low-dimensional single-trial latent factors [33], has also been used to predict behavioral responses during learning [34–36] and more generally on a trial-by-trial basis [34,37–42]. Further, low dimensional trajectories might be computationally desirable, as they are robust to noise [43], and may support both memory [44–46] and generalization [45].

These approaches outlined above have led to new ways of conceptualizing computations performed by neural populations, but they are not a panacea. Descriptions of population dynamics often occur in a latent space abstracted from the underlying neurons [8]. Thus population-computation hypotheses can be both hard to intuitively understand, and direct tests of these hypotheses would require arbitrary spatiotemporal control of neural dynamics [47,48]. Further, these approaches often include dimensionality reduction, a technique which can be sensitive to details like behavioral complexity, number of neurons recorded, fraction of a population encoding the variable of interest, length of recording, amount of preprocessing, and many other factors [49,50].

Carving neural representations at their joint: Identifying functional neuron types

The low-dimensional dynamics perspective is often motivated by the long-standing challenge of distilling the diverse neural tuning profiles found in frontal cortex to a principled understanding of computations. As an alternative to considering the instantaneous neural activity as a point in an N-neuron dimensional space, each neuron can be considered as a point in a space defined by its tuning to different task and behavioral features, as well

as by its temporal activity profile across the trial. A handful of recent studies have examined the structure in this high dimensional space and have discovered that neuronal profiles are not uniformly/randomly distributed, rather they cluster together in dynamics and/or tuning. Under the hypothesis that neurons in these groups play similar roles in population computations, these studies provide a new link between neural identity and computational role. Thus, these studies, reviewed in this section, highlight a new perspective on cell-types, definable by functional properties (e.g., temporal dynamics, high dimensional tuning).

Analyzing neural activity during decisions that required integrating sensory and reward information, Hirokawa et al. [51] found that lateral orbitofrontal cortical (lOFC) neurons were tuned to privileged mixtures of task features which corresponded to specific "decision variables." They developed a clustering approach to identify neurons with similar response profiles, in terms of tuning or temporal dynamics, and found thatthe neural population was highly structured with distinct groups of neurons having different coding profiles. Surprisingly each of the functional types had a clear interpretation based on a simple model of the choice behavior, quantitatively matching decision variables such as decision confidence, reward magnitude, and integrated value (Figure 2).

In search of functional clusters, Hocker et al. [55] re-analyzed recording data from a value-based decision task that separately manipulated reward probability and amount to understand the role of reward history under risk [56]. They found multiple functional clusters of lateral OFC neurons with one cluster encoding reward history directly preceding the animal's choice. similar to a functional cluster observed by Hirokawa et al [51] The authors suggest this group of neurons could play a causal role in driving reward-history behavior biases.

Namboodiri et al.[52] applied the same clustering approach to calcium imaging data, so they could investigate the activity of ventral orbital and medial OFC (vmOFC) neurons while mice learned a Pavlovian trace-conditioning task. Importantly, the different vmOFC clusters had distinct temporal (and tuning) dynamics that peaked at different phases of the task, allowing the authors to use optogenetics to specifically disrupt activity in only subsets of clusters, to tease apart differential cluster contributions to association learning, maintenance, and extinction. Finally, Onken et al. developed a statistical framework to establish the existence of categorical groupings of neurons in primate OFC [53]. Interestingly, in many cases mentioned here, the authors found a correspondence between clusters found in a space only consisting of tuning features, and one consisting of only neural dynamics. That is, neurons code for specific conjunctions of features, at specific phases intrial progression, as would be expected if the neural activity represented the dynamical evolution of a cognitive algorithm.

Together, these studies suggest that neural activity in frontal cortex might have interpretable, categorical structure along specific temporal and tuning dimensions. If this is true, why have previous studies reported randomly mixed neural activity? This discrepancy may be due to differences in brain regions or species analyzed; most studies discussed in this section assessed neural coding in orbitofrontal cortex, whereas random mixed selectivity has been reported in prefrontal cortex [23] or even non-frontal regions like posterior parietal cortex

[53]. Alternatively, it is possible that what seems like random mixed selectivity might simply be a function of behavioral tasks that are either too low-dimensional, or that are ill-matched to the brain region studied [51].

To investigate how task demands might constrain neural coding strategies, Dubreuil et al. [54] analyzed recurrent neural networks trained to perform a variety of common neuroscience tasks. By analyzing their trained networks, they determined that the presence of specialized groups of neurons is necessary when task demands require flexible computations. Similarly, by analyzing neural dynamics in recurrent neural networks, monkeys, and in humans (using functional magnetic resonance imaging), all trained to perform a context dependent decision-making task, Flesch et al. [55] found more evidence for categorical computations than for random high-dimensional projections.

Thus to observe categorical encoding it seems critical to use appropriately cognitively demanding tasks that provide and require a rich embedding of the task-relevant variables.

We emphasize that these theoretical results do not simply apply to the task the animal is performing, but rather to the computations required of *the brain region of interest*. For example, in the case of Hirokawa et al., the behavioral task required that animals form a decision integrating time-varying reward expectations with sensory uncertainty, a flexible computation with multiple components that critically depends on lOFC [56]. To drive home this point, as a counterexample we would not expect that neurons in visual cortex would display categorical encoding of these specific decision variables, even in an identical task.

In summary, if one wishes to adjudicate between categorical and randomly mixed coding types, it 'is important that the task provides a sufficient number of coding dimensions and that those dimensions fall within the brain region's computational roles. Thus, given the important role the frontal cortex plays in flexible, context-dependent decision-making, these theoretical and experimental results support the idea that distinct cognitive operations might be realized by distinct groups of frontal neurons each with specific algorithmic roles. Can we find a correspondence between these functional neuron types and the more well-known structural neuron types?

Cell-type-specific computations in frontal cortex

The last decade has seen a revolution in our toolkit for categorizing the taxonomy of cortical neurons, based on gene expression [57], projection target [58], and single-cell morphology [59]. These advances have led to the identification of specific functional roles of different cortical inhibitory neuron types, such as a gain modulation or disinhibition [60,61]. There is also increasing appreciation for a division of labor across inhibitory neuron types during behavior with a potential to control the flow of information across regions at behavioral time scales ([62–67]) Similarly, new studies have demonstrated the organization of long-range subcortical connectivity of frontal pyramidal neurons [68–72] into motifs, often suggestive of an operational role. For example, a projection from the prefrontal cortex to the basal forebrain might be suggestive of top-down control of arousal [73].

In a follow-up experiment to the functional clustering work discussed above, Hirokawa et al. [51] used electrophysiology and optogenetic tagging to record from a subset of lOFC neurons that project to the striatum (Figure 3a–c). These neurons were astonishingly uniform in their activity profile, with sustained activity persisting from reward receipt through the inter-trial interval, and only diminishing with the onset of the next trial's stimuli. These neurons all belonged to one of two functional clusters that encoded positive and negative integrated value, respectively. Intriguingly, using a different behavioral task, Hocker et al. [74] identified a functional cluster of OFC neurons whose dynamics bore a striking resemblance to the striatumprojecting OFC neurons in Hirokawa et al. (Figure 3d). These neural responses show key features required for a temporal credit assignment mechanism critical in reinforcement learning [75]. In two entirely different tasks, Terra et al. [76] and Bari et al. [77] found the medial PFC-striatum projecting neurons also displayed sustained activity across task phases. Taken together, these results reveal that the frontal cortex to striatum projection comprises a core circuit module for prolonged communication of decision variables to the behavioral learning and execution circuitry of the striatum.

The cell-type-specific architecture of cortical dynamics has been extensively studied in a remarkable series of experiments by Svoboda and colleagues in motor cortex. Guo et al [78] analyzed a subset of layer 5b neurons that project to the thalamus, which in turn project reciprocally back to the same local cortical region. Such recurrent loops are a ubiquitous feature in cortical architecture [71], and thalamocortical loops in particular are crucial for controlling cortical excitability – runaway activity in these loops can lead to epileptic seizures [79]. Indeed, the authors demonstrated that these thalamocortical loops in mice were responsible for sustaining delay period preparatory motor activity, in anterolateral motor cortex (ALM) during a delayed response task. Economo et al. [80] further elucidated the role of different subgroups of pyramidal projection neurons–while a subset of neurons in the upper part of layer 5b neurons project to the thalamus, a distinct subgroup of pyramidal neurons in the lower part of layer 5b project to the medulla, a motor output nuclei in the brainstem. They demonstrated that these anatomical projections are the circuit mechanism through which ALM pyramidal neurons were able to segregate preparatory motor activity and activity that drives motor output (Figure 3e–g). These results provide a cell-type-specific solution for a long-standing mystery in motor neuroscience: how does preparatory activity in motor cortex exist without driving motor output? A similar role for thalamus projecting medial PFC neurons in sustained activity has been identified by two different groups, Bolkan et al. [81] and Schmitt et al. [82], pointing toa cell-type-specific computational motif, in which thalamocortical loopssupport sustained activity.

It is important to note that neural activity is not always found to be separable along axes defined by projection-output specified cell types. For example, Spellman et al.[83] find heterogeneous coding of specific task-features in PFC-striatum projection neurons (however, they did find a laminar gradient of reward coding in mouse PFC, following the distribution of input axons from the anterior cingulate cortex). Negative results in projection-specific coding studies could arise for several reasons, including the lack of a specific need for that projection in the task at hand, or the dominance of another projection which anatomically overlaps with the studied projection. Alternately neural coding in a given pathway might seem "mixed" when considered with respect to particular experimental variables of interest–

however, there might still be patterns in the tuning to all task features (especially given high enough task complexity), and in their dynamical time course. This less biased search for patterns in neural activity reflects a search for an algorithmic role for neural activity (e.g., the maintenance of sustained activity, which in turn is a mechanism of working memory), rather than the correlation of neural activity with experimenter pre-supposed task features.

Thus, we refer to "functional cell-types" in cases when a group of neurons is found to play a shared computational role. Functional cell-types need not be connectivity defined or genetically distinct, however given that connectivity and biophysical properties determine neural dynamics, it is our belief that in most cases "functional" cell-types will overlap with those defined biologically. Further elucidation of the prevalence of functional types will be necessary to determine whether these types should be considered as computational primitives necessary for flexible cognitive behavior, as predicted by theory [54].

Future & outlook

Broadly speaking, our hypothesis is that cognition is constructed from building blocks of certain core dynamical computations, with distinct contributions from physiologically different types of neurons. The Lego set of cognition may include several of these cell-type-specified building blocks, as well as general learning-based modules whose activity is only understandable at a (sub)population level. We hope the studies we have reviewed here impress the reader that despite the many advantages of population approaches, there is often practical utility in anatomically grounded approaches. We see no reason that just because some neurons in a particular region or task display heterogeneous, uninterpretable single-neuron coding, that the entire program of identifying individual neurons with interpretable roles should be abandoned.

Frontal cortex, like the rest of neocortex, has stereotyped microcircuitry with specific long-range input and output connections. Theoretically, such network connectivity is known to determine individual neuron dynamics [84–86] and subcortical output targets are often tied to specific sensors and effectors (e.g., the retina, but also hormone release, etc.), strongly constraining their function. From an evolutionary perspective, it seems unparsimonious that biology would go to the trouble of substantial entropy reduction to create and genetically encode the exquisite order of brain circuit architecture [71,87], if the repeated long-range and local circuit modules did not play some adaptive role. Indeed, there is increasing understanding about evolutionary conservation of key long-range projections and other cell-types in vertebrates [88–90] that may be constrained by shared ecological objective functions [91].

We propose that biological cell-types with developmentally specified architecture and physiology can be thought of as an evolutionary prior on the type of computation performable by a given set of neurons. The mapping between computation and biological cell-types is bound to arise as evolution retains architectures that provide useful constraints on neural dynamics across ecologically relevant behaviors. As a practical consequence, observing the activity of specific cell-types can serve the same role as dimensionality reduction across anonymous neural populations–providing an interpretable handle on computations.

Our hypothesis can be seen as a marriage between the "Hopfieldian" (neural dynamics are the primary level of explanation for cognition) and the "Sherringtonian" (single neurons and their interactions are the primary explanation for cognition) views. As argued by Barack and Krakauer [92], we propose that neural dynamics are indeed the appropriate level at which to search for encoding of cognitive variables. However, unlike Barack and Krakauer, we believe these dynamics can be (and indeed must be, in cases when task complexity is high and connectivity is non-random [54,84]) implemented by functional classes of neurons, corresponding to groups of neurons with specific non-random connectivity motifs. If we are correct, the field's increasing knowledge of cell-type-specific neural circuitry will provide badly needed prior information in the search for what computations are implemented by neural populations—an under-constrained problem in any realistic experimental regime.

Thus, rather than be disheartened by the need to include more biological detail into our models, we are excited about the opportunity those elements provide in progress towards understanding the frontal cortex and beyond. Not only will such a biologically grounded understanding of neural computations help us validate and generalize our theories, but it holds the potential for new understanding of neuropsychiatric treatments—-that typically act, through drugs, at the level of neurons and the receptors they express, and only indirectly influence population dynamics.

Acknowledgments

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

1. Fuster JM: The prefrontal cortex–an update: time is of the essence. Neuron 2001, 30:319–333. [PubMed: 11394996]

2. Miller EK: The prefrontal cortex and cognitive control. Nat Rev Neurosci 2000, 1:59–65. [PubMed: 11252769]

3. Bicks LK, Koike H, Akbarian S, Morishita H: Prefrontal cortex and social cognition in mouse and man. Front Psychol 2015, 6: 1805. [PubMed: 26635701]

4. Fleming SM, Dolan RJ: The neural basis of metacognitive ability. Philos. Trans. R. Soc. B Biol. Sci 2012, 367: 1338–1349.

5. Lee TS: Computations in the early visual cortex. J Physiol Paris 2003, 97:121–139. [PubMed: 14766138]

6. Krizhevsky A, Sutskever I, Hinton GE: ImageNet classification with deep convolutional neural networks. in Advances in neural information processing systems, 25. Curran Associates, Inc.; 2012.

7. Yamins DLK, et al. : Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci USA 2014, 111:8619–8624. [PubMed: 24812127]

8. Duncker L, Sahani M: Dynamics on the manifold: identifying computational dynamical activity from neural population recordings. Curr Opin Neurobiol 2021, 70:163–170. [PubMed: 34837752]

9. Vyas S, Golub MD, Sussillo D, Shenoy KV: Computation through neural population dynamics. Annu Rev Neurosci 2020, 43:249–275. [PubMed: 32640928]

10. Jazayeri M, Ostojic S: Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. Curr Opin Neurobiol 2021, 70:113–120. [PubMed: 34537579]

11. Fishell G, Kepecs A: Interneuron types as attractors and controllers. Annu Rev Neurosci 2020, 43:1–30. [PubMed: 31299170]

12. Fusi S, Miller EK, Rigotti M: Why neurons mix: high dimensionality for higher cognition. Curr Opin Neurobiol 2016, 37: 66–74. [PubMed: 26851755]

13. Saxena S, Cunningham JP: Towards the neural population doctrine. Curr Opin Neurobiol 2019, 55:103–111. [PubMed: 30877963]

14. Mante V, Sussillo D, Shenoy KV, Newsome WT: Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 2013, 503:78–84. [PubMed: 24201281]

15. Raposo D, Kaufman MT, Churchland AK: A category-free neural population supports evolving demands during decision-making. Nat Neurosci 2014, 17:1784–1792. [PubMed: 25383902]

16. Cunningham JP, Yu BM: Dimensionality reduction for large-scale neural recordings. Nat Neurosci 2014, 17:1500–1509. [PubMed: 25151264]

17. Chung S, Abbott LF: Neural population geometry: an approach for understanding biological and artificial neural networks. Curr Opin Neurobiol 2021, 70:137–144. [PubMed: 34801787]

18. Giusti C, Ghrist R, Bassett DS: Two's company, three (or more) is a simplex: algebraic-topological tools for understanding higher-order structure in neural data. ArXiv160101704 Math Q-Bio; 2016.

19. BrainGate - Turning Thought Into Action. https://www.braingate.org/.

20. Sadtler PT, et al. : Neural constraints on learning. Nature 2014, 512:423–426. [PubMed: 25164754]

21. Kaufman MT, Churchland MM, Ryu SI, Shenoy KV: Cortical activity in the null space: permitting preparation without movement. Nat Neurosci 2014, 17:440–448. [PubMed: 24487233]

22. Gallego JA, Perich MG, Miller LE, Solla SA: Neural manifolds for the control of movement. Neuron 2017, 94:978–984. [PubMed: 28595054]

23. Pandarinath C, et al. : Latent factors and dynamics in motor cortex and their application to brain–machine interfaces. J Neurosci 2018, 38:9390–9401. [PubMed: 30381431]

24. Aoi MC, Mante V, Pillow JW: Prefrontal cortex exhibits multi-dimensional dynamic encoding during decision-making. Nat Neurosci 2020, 23:1410–1420. [PubMed: 33020653] **The authors show that prefrontal cortex dynamics during context dependent decision-making display time-evolving dynamics, with both relevant and irrelevant sensory information maintained until choice by a multi-dimensional population representation.

25. Sohn H, Narain D, Meirhaeghe N, Jazayeri M: Bayesian computation through cortical latent dynamics. Neuron 2019, 103:934–947.e5. [PubMed: 31320220] The authors show that prior beliefs about interval timing duration temporally "warp" population population trajectories during decision making. This provides a population-based mechanism for Bayesian integration of evidence and prior beliefs.

26. Ebitz RB, Tu JC, Hayden BY: Rules warp feature encoding in decision-making circuits. PLoS Biol 2020, 18, e3000951. The authors found that following different stimulus-response rules warped the geometry of neural representations in orbitofrontal cortex and striatum, expanding along the rule-relevant dimension and shrinking along the rule-irrelevant one.

27. Malagon-Vina H, Ciocchi S, Passecker J, Dorffner G, Klausberger T: Fluid network dynamics in the prefrontal cortex during multiple strategy switching. Nat Commun 2018, 9:309. [PubMed: 29358717]

28. Remington ED, Egger SW, Narain D, Wang J, Jazayeri M: A dynamical systems perspective on flexible motor timing. Trends Cognit Sci 2018, 22:938–952. [PubMed: 30266152]

29. Jazayeri M, Shadlen MN: A neural mechanism for sensing and reproducing a time interval. Curr. Biol. CB 2015, 25: 2599–2609. [PubMed: 26455307]

30. Remington ED, Narain D, Hosseini EA, Jazayeri M: Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. Neuron 2018, 98:1005–1019.e5. [PubMed: 29879384] The authors provide evidence that neural trajectories in dorsomedial frontal cortex in a timing task could be best explained as dynamical system in which task demands reconfigure the systems initial conditions and input.

31. Wang J, Narain D, Hosseini EA, Jazayeri M: Flexible timing by temporal scaling of cortical responses. Nat Neurosci 2018, 21: 102–110. [PubMed: 29203897] The authors study the flexibility

of behavioral speed control and identify a temporal invariance of firing rate dynamics that can be explain by recurrent network dynamics. Temporal scaling can emerge naturally in recurrent networks with scaling controlled by external input.

32. Egger SW, Remington ED, Chang C-J, Jazayeri M: Internal models of sensorimotor integration regulate cortical dynamics. Nat Neurosci 2019, 22:1871–1882. [PubMed: 31591558]

33. Williams AH, Linderman SW: Statistical neuroscience in the single trial limit. ArXiv210305075 Q-Bio Stat; 2021.

34. Bartolo R, Averbeck BB: Prefrontal cortex predicts state switches during reversal learning. Neuron 2020, 106: 1044–1054.e4. [PubMed: 32315603]

35. Bartolo R, Saunders RC, Mitz AR, Averbeck BB: Dimensionality, information and learning in prefrontal cortex. PLoS Comput Biol 2020, 16, e1007514.

36. Cohen Y, Schneidman E, Paz R: The geometry of neuronal representations during rule learning reveals complementary roles of cingulate cortex and putamen. Neuron 2021, 109: 839–851.e9. [PubMed: 33484641]

37. Inagaki HK, Fontolan L, Romani S, Svoboda K: Discrete attractor dynamics underlies persistent activity in the frontal cortex. Nature 2019, 566:212–217. [PubMed: 30728503]

38. Kurikawa T, Haga T, Handa T, Harukuni R, Fukai T: Neuronal stability in medial frontal cortex sets individual variability in decision-making. Nat Neurosci 2018, 21:1764–1773. [PubMed: 30420732]

39. Li N, Daie K, Svoboda K, Druckmann S: Robust neuronal dynamics in premotor cortex during motor planning. Nature 2016, 532:459–464. [PubMed: 27074502]

40. Pandarinath C, et al. : Inferring single-trial neural population dynamics using sequential auto-encoders. Nat Methods 2018, 15:805–815. [PubMed: 30224673]

41. Peron S, et al. : Recurrent interactions in local cortical circuits. Nature 2020, 579:256–259. [PubMed: 32132709] * The authors use targeted photoablation to test the role of individual cortical neurons in distributed representations. They show that the selective degradation of representations similar to the ablated neurons can be explained with recurrent networks producing input-specific amplification.

42. McGinty VB, Lupkin SM: Value signals in orbitofrontal cortex predict economic decisions on a trial-to-trial basis. 2021, 434452, 10.1101/2021.03.11.434452. 03.11 , https://www.biorxiv.org/content/10.1101/2021.03.11.434452v1.

43. Bernardi S, et al. : The geometry of abstraction in the Hippo-campus and prefrontal cortex. Cell 2020, 183:954–967.e21. [PubMed: 33058757] *The authors found behavioral generalization was supported by particular geometries of complex neural representation in hippocampus and prefrontal cortex, changing with task events but invariant to linear readout.

44. Cavanagh SE, Towers JP, Wallis JD, Hunt LT, Kennerley SW: Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. Nat Commun 2018, 9: 3498. [PubMed: 30158519]

45. Cueva CJ, et al. : Low-dimensional dynamics for working memory and time encoding. Proc Natl Acad Sci USA 2020, 117: 23021–23032. [PubMed: 32859756]

46. Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC: Correlated variability modifies working memory fidelity in primate pre-frontal neuronal ensembles. Proc Natl Acad Sci USA 2017, 114: E2494–E2503. [PubMed: 28275096]

47. Jazayeri M, Afraz A: Navigating the neural space in search of the neural code. Neuron 2017, 93:1003–1014. [PubMed: 28279349]

48. Shenoy KV, Kao JC Measurement: Manipulation and modeling of brain-wide neural population dynamics. Nat Commun 2021, 12:633. [PubMed: 33504773]

49. Gao P, et al.: A theory of multineuronal dimensionality, dynamics and measurement. 2017:214262, 10.1101/214262. https://www.biorxiv.org/content/10.1101/214262v2.

50. Engel TA, Steinmetz NA: New perspectives on dimensionality and variability from large-scale cortical dynamics. Curr Opin Neurobiol 2019, 58:181–190. [PubMed: 31585331]

51. Hirokawa J, Vaughan A, Masset P, Ott T, Kepecs A: Frontal cortex neuron types categorically encode single decision variables. Nature 2019, 576:446–451. [PubMed: 31801999] ** Hirokawa et al. demonstrate that neural activity recorded in rat OFC during a value-guided sensory decision-

making task can be categorized into distinct functional clusters, and neural activity in each of these clusters can be related to canonical "decision variables" – components of a normative model of the animals' behavior. Finally, the authors record from a connectivity- defined cell types, OFC neurons projecting to the striatum, and demonstrate that the recorded neurons carry to a single decision variable.

52. Namboodiri VMK, et al. : Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. Nat Neurosci 2019, 22:1110–1121. [PubMed: 31160741] ** Namboodiri et al. record activity in mouse OFC and report that single neurons can be segregated into distinct functional clusters. They then record from subpopulations of OFC neurons that project to the VTA and demonstrated that only a particular subsets of OFC activity was relayed by VTA projection neurons.

53. Onken A, Xie J, Panzeri S, Padoa-Schioppa C: Categorical encoding of decision variables in orbitofrontal cortex. PLoS Comput Biol 2019, 15, e1006667. *The authors identify categorical, non-mixed, representations of decision variables such as offer value and choice outcome in orbitofrontal cortex during economic decisions.

54. Dubreuil A, Valente A, Beiran M, Mastrogiuseppe F, Ostojic S: The role of population structure in computations through neural dynamics. Nat Neurosci 2022, 25:783–794. [PubMed: 35668174] ** The authors find that neural networks trained to perform flexible, context-dependent sensory decision-making tasks rely on categorical groups of neurons with different connectivity and coding patterns, while they find that networks trained on simpler tasks do not require these separate groups of neurons, and instead rely on connectivity and neural activity that appears approximately gaussian.

55. Flesch T, Juechems K, Dumbalska T, Saxe A, Summerfield C: Orthogonal representations for robust context-dependent task performance in brains and neural networks. Neuron 2022, 110:1258–1270.e11. [PubMed: 35085492] ** The authors analyze simple recurrent neural networks trained on context dependent decision-making tasks. Using either low-variance, or high-variance weight initializations, they demonstrate that RNNs can solve these tasks using either a "rich" categorical type encoding, or a "lazy" high dimensional (and mixed-selectivity) coding strategy. Using their RNN models they make distinct predictions about features of population activity they expect under each respective coding strategy and demonstrate that neural activity in monkey PFC and human frontopariatel networks resembled the "rich" regime rather than the lazy regime.

56. Walton ME, Behrens TEJ, Noonan MP, Rushworth MFS: Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. Ann N Y Acad Sci 2011, 1239:14–24. [PubMed: 22145871]

57. Callaway EM, et al. : A multimodal cell census and atlas of the mammalian primary motor cortex. Nature 2021, 598:86–102. [PubMed: 34616075]

58. Kebschull JM, et al. : High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. Neuron 2016, 91:975–987. [PubMed: 27545715]

59. Economo MN, et al. : A platform for brain-wide imaging and reconstruction of individual neurons. Elife 2016, 5, e10566.

60. Geiller T, et al. : Large-Scale 3D two-photon imaging of molecularly identified CA1 interneuron dynamics in behaving mice. Neuron 2020, 108:968–983.e9. [PubMed: 33022227]

61. Letzkus JJ, Wolff SBE, Lüthi A: Disinhibition, a circuit mechanism for associative learning and memory. Neuron 2015, 88: 264–276. [PubMed: 26494276]

62. Hangya B, Pi H-J, Kvitsiani D, Ranade SP, Kepecs A: From circuit motifs to computations: mapping the behavioral repertoire of cortical interneurons. Curr Opin Neurobiol 2014, 26:117–124. [PubMed: 24508565]

63. Kvitsiani D, et al. : Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature 2013, 498:363–366. [PubMed: 23708967]

64. Pi H-J, et al. : Cortical interneurons that specialize in disinhibitory control. Nature 2013, 503:521–524. [PubMed: 24097352]

65. Donato F, Rompani SB, Caroni P: Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. Nature 2013, 504:272–276. [PubMed: 24336286]

66. Pinto L, Dan Y: Cell-type-specific activity in prefrontal cortex during goal-directed behavior. Neuron 2015, 87:437–450. [PubMed: 26143660]

67. Dudok B, et al. : Recruitment and inhibitory action of hippocampal axo-axonic cells during behavior. Neuron 2021, 109: 3838–3850.e8. [PubMed: 34648750]

68. Vander Weele CM, et al. : Dopamine enhances signal-to-noise ratio in cortical-brainstem encoding of aversive stimuli. Nature 2018, 563:397–401. [PubMed: 30405240]

69. Otis JM, et al. : Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. Nature 2017, 543: 103–107. [PubMed: 28225752]

70. Lui JH, et al. : Differential encoding in prefrontal cortex projection neuron classes across cognitive tasks. Cell 2021, 184: 489–506.e26. [PubMed: 33338423] * The authors find a one-to-one mapping between transcriptomic and projection neuron types in prefrontal cortex but heterogeneous task encoding. Prefrontal cortex to PAG projection neurons preferentially encode choice while contralateral projections tend to encode reward context.

71. Luo L: Architectures of neuronal circuits. Science 2021, 373, eabg7285.

72. Warden MR, et al. : A prefrontal cortex–brainstem neuronal projection that controls response to behavioural challenge. Nature 2012, 492:428–432. [PubMed: 23160494]

73. Demeter E, Sarter M: 3.25 - ascending systems – top down control: noradrenergic and cholinergic control of attention and learning. In Byrne JH, 463–473. Academic Press; 2017, 10.1016/B978-0-12-809324-5.21084-5.

74. Hocker DL, Brody CD, Savin C, Constantinople CM: Sub-populations of neurons in lOFC encode previous and current rewards at time of choice. Elife 2021, 10, e70129. **Hocker et al. analyze data from rat OFC and find neural activity can be segregated into distinct functional clusters. They identify a cluster whose activity is sustained throughout the ITI and represents reward history. The activity of neurons in this cluster b ore a striking resemblence to the activty of the striatum projecting neurons identified in Hirokawa et al., 2019.

75. Chau BKH, et al. : Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. Neuron 2015, 87:1106–1118. [PubMed: 26335649]

76. Terra H, et al. : Prefrontal cortical projection neurons targeting * dorsomedial striatum control behavioral inhibition. Curr Biol 2020, 30:4188–4200.e5. [PubMed: 32888489] * The authors show that dorsomedial prefrontal cortex neurons that project to the striatum show selectivity firing rate dynamics, persistent activity during behavioral inhibitory control. Turning off this projection impairs behavioral inhibition.

77. Bari BA, et al. : Stable representations of decision variables for flexible behavior. Neuron 2019, 103:922–933.e7. [PubMed: 31280924]

78. Guo ZV, et al. : Maintenance of persistent activity in a frontal thalamocortical loop. Nature 2017, 545:181–186. [PubMed: 28467817] ** Guo et al. perform optogenetic manipulations and recordings from both mouse motor cortex andmotor thalamus and determine that the recurrent cortex-thalamus loop is necessary for the maintenance of persistent activity during the delay phase of a delayed response task.

79. McCormick DA, Contreras D: On the cellular and network bases of epileptic seizures. Annu Rev Physiol 2001, 63:815–846. [PubMed: 11181977] ** Economo et al. perform recordings and manipulations of two subpopulations of motor cortex layer 5b neurons, one projecting to the medulla and the other to the thalamus. They determine that the medulla-targeting subpopulation of layer 5b neurons carry action specific information directly before movement onset, while the thalamic projecting subpopulation maintains choice specific information throughout the delay period in a delayed response task.

80. Economo MN, et al. : Distinct descending motor cortex pathways and their roles in movement. Nature 2018, 563: 79–84. [PubMed: 30382200]

81. Bolkan SS, et al. : Thalamic projections sustain prefrontal activity during working memory maintenance. Nat Neurosci 2017, 20:987–996.* [PubMed: 28481349]

82. Schmitt LI, et al. : Thalamic amplification of cortical connectivity sustains attentional control. Nature 2017, 545:219–223. [PubMed: 28467827] * These two papers along with Guo and colleagues above show that persistent activity in frontal cortex and motor cortex, respectively, is maintained via a cortico-thalamocortical loop.

83. Spellman T, Svei M, Kaminsky J, Manzano-Nieves G, Liston C: Prefrontal deep projection neurons enable cognitive flexibility via persistent feedback monitoring. Cell 2021, 184:2750–2766.e17. [PubMed: 33861951] * The authors find that functional responses of prefrontal projection neurons do not vary with projection target but rather follow a topo-graphic gradient, which matches the pattern of specific inputs to PFC.

84. Mastrogiuseppe F, Ostojic S: Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. Neuron 2018, 99:609–623.e29. [PubMed: 30057201] *The authors study recurrent networks with minimally structured connectivity and identify testable experimental predictions for the relationship between connectivity, low-dimensional dynamics, and computations represented

85. Recanatesi S, Ocker GK, Buice MA, Shea-Brown E: Dimensionality in recurrent spiking networks: global trends in activity and local origins in connectivity. PLoS Comput Biol 2019, 15, e1006446.

86. Litwin-Kumar A, Doiron B: Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat Neurosci 2012, 15:1498–1505. [PubMed: 23001062]

87. Zador AM: A critique of pure learning and what artificial neural networks can learn from animal brains. Nat Commun 2019, 10: 3770. [PubMed: 31434893]

88. Tosches MA, Laurent G: Evolution of neuronal identity in the cerebral cortex. Curr Opin Neurobiol 2019, 56:199–208. [PubMed: 31103814]

89. Pessoa L, Medina L, Hof PR, Desfilis E: Neural architecture of the vertebrate brain: implications for the interaction between emotion and cognition. Neurosci Biobehav Rev 2019, 107: 296–312. [PubMed: 31541638]

90. Karten HJ: Evolutionary developmental biology meets the brain: the origins of mammalian cortex. Proc Natl Acad Sci USA 1997, 94:2800–2804. [PubMed: 9096300]

91. Hasson U, Nastase SA, Goldstein A: Direct fit to nature: an evolutionary perspective on biological and artificial neural networks. Neuron 2020, 105:416–434. [PubMed: 32027833]

92. Barack DL, Krakauer JW: Two views on the cognitive brain. Nat Rev Neurosci 2021, 22:359–371. [PubMed: 33859408]
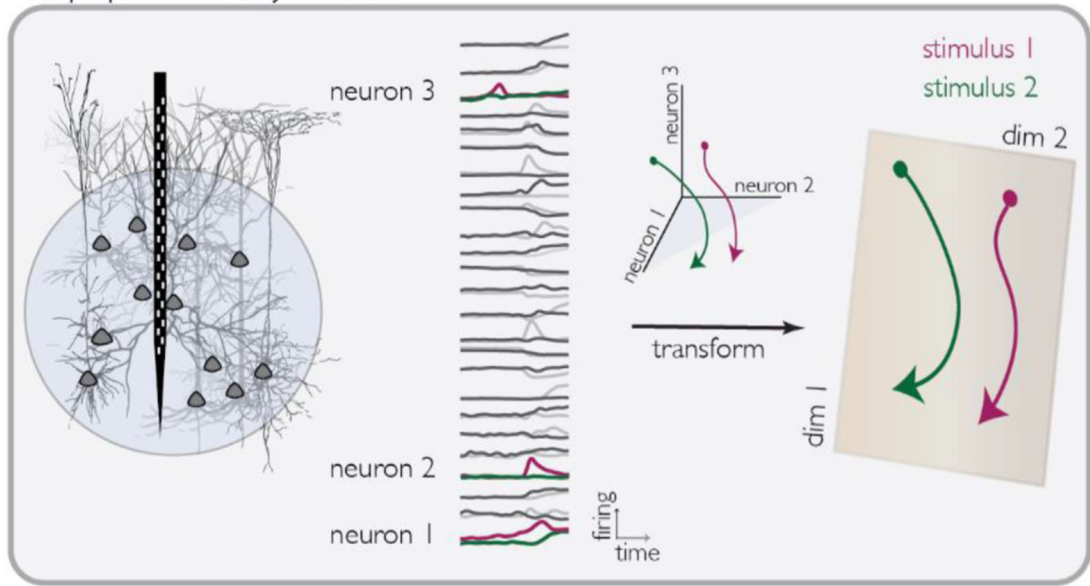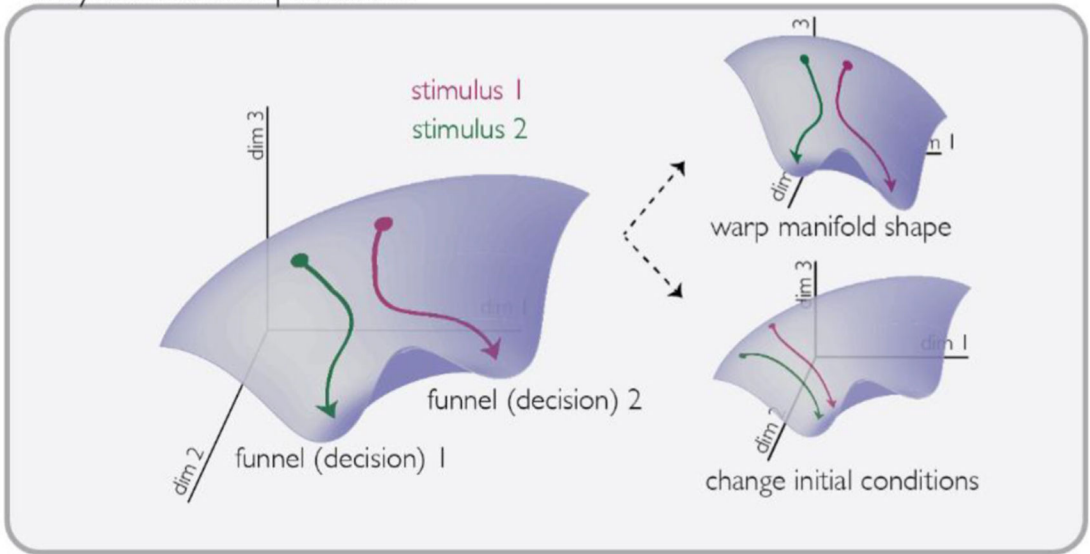
Figure 1.
a. Schematic of population analysis of neural activity, showing transformation from recorded PSTHs to neural state space, to a dimensionality reduced space. **b**. Schematic of changes in population activity manifolds observed in flexible context-dependent decision-making tasks. Insets show two ways (warping of manifold shape, and changing of initial condition) that population activity can be updated based on task context.
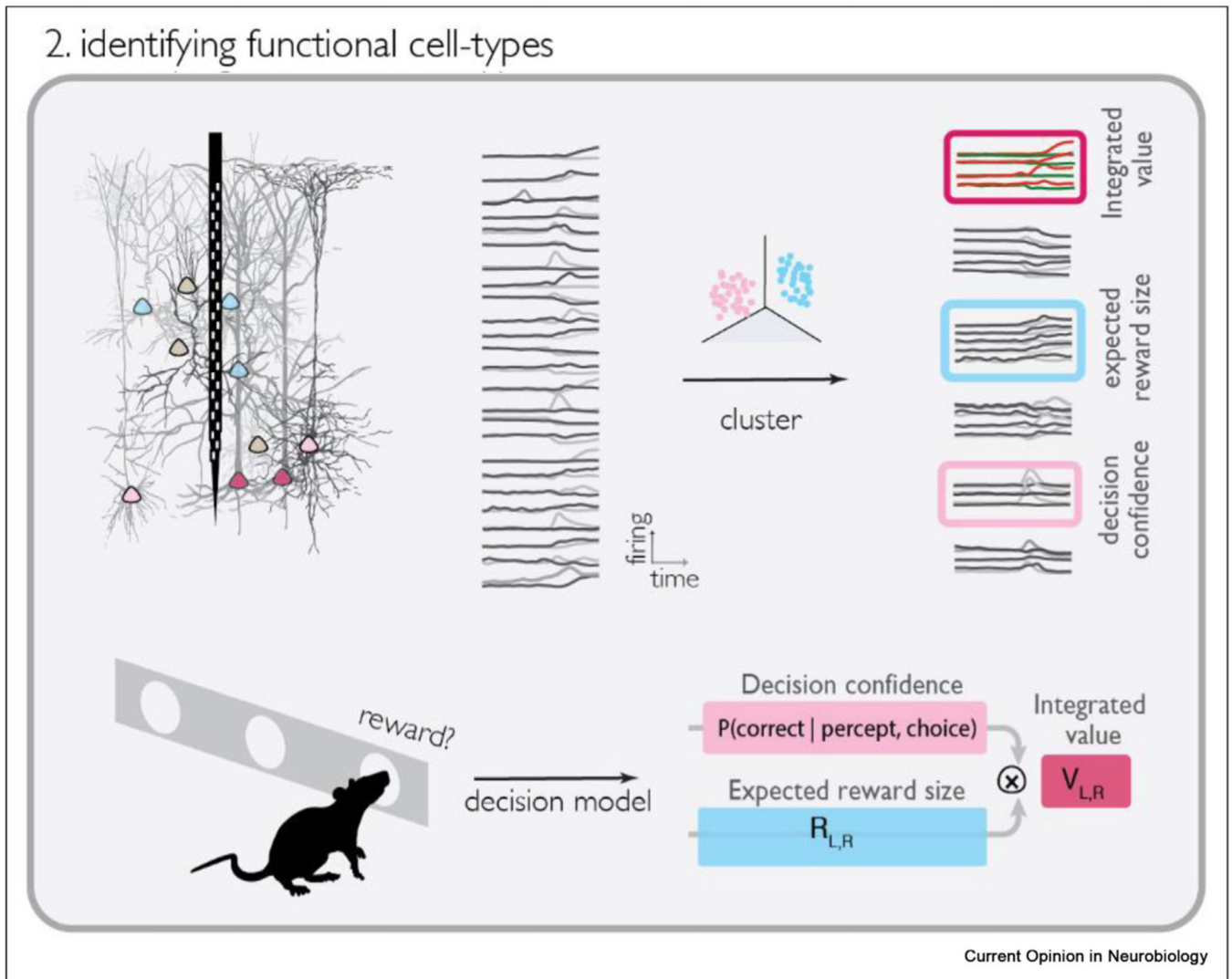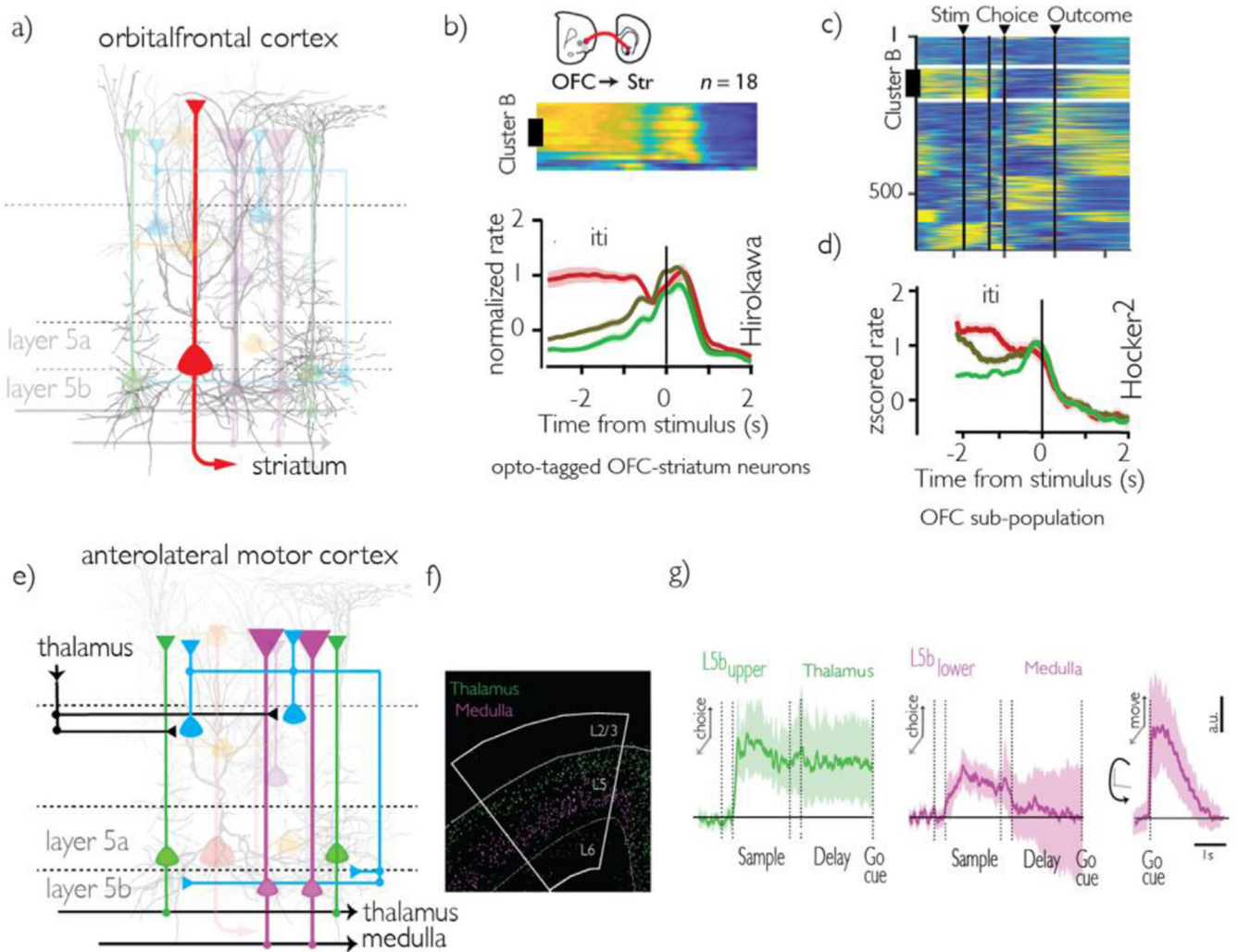
Figure 2.
Schematic of functional clustering approach. The top row depicts clustering neural activity according to tuning or dynamics, then identifying the clusters as decision variables in a model of rodent behavior (bottom row).

Figure 3.

a. Schematic of OFC-striatum projection. **b**. Recordings from a subpopulation of OFC neurons that project to the striatum (identified via optogenetic tagging), whose activity is sustained throughout the inter-trial-interval and represents integrated value of the previous choice. Figure reproduced from [51]. **c**. Clustered population raster of all recorded OFC neurons. The population presented in panel b. closely resembles the activity of the cluster marked "B." Figure reproduced from [51]. d. Activity in a single cluster of OFC neurons whose activity is sustained throughout the inter-trial-interval and represents integrated value of the previous choice, and whose activity resembles that of the striatal-projecting neurons in panels b. Figure reproduced from [74]. **e**. Schematic of ALM circuits identified in a series of papers by the Svoboda laboratory. **f**. Histology showing the segregation of medulla and thalamus projecting neurons in layer 5 of ALM. Figure reproduced from [80]. **g**. (Left) choice selectivity in thalamus projecting and medulla projecting ALM neurons respectively throughout the sample and delay task periods. Only thalamus projecting neurons maintain

choice selectivity during the delay period. (Right) medulla projecting neurons display movement selectivity after the go cue.