

Theoretical principles for illuminating sensorimotor processing with brain-wide neuronal recordings

Tirthabir Biswas, William E. Bishop, James E. Fitzgerald

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA.

Abstract

Modern recording techniques now permit brain-wide sensorimotor circuits to be observed at single-neuron resolution in small animals. Extracting theoretical understanding from these recordings requires principles that organize findings and guide future experiments. Here we review theoretical principles that shed light onto brain-wide sensorimotor processing. We begin with an analogy that conceptualizes principles as streetlamps that illuminate the empirical terrain, and we illustrate the analogy by showing how two familiar principles apply in new ways to brain-wide phenomena. We then focus the bulk of the review on describing three more principles that have wide utility for mapping brain-wide neural activity, making testable predictions from highly parameterized mechanistic models, and investigating the computational determinants of neuronal response patterns across the brain.

Multiple principles will be needed to understand sensorimotor processing

Understanding brain-wide circuit mechanisms of behavior is a central goal of neuroscience. Sensorimotor circuits require multiple processing stages to extract relevant sensory information, select a course of action, and generate appropriate motor commands. Considerable effort has already gone into deciphering individual stations within such circuits, but understanding integrated sensorimotor processing remains challenging. In particular, sensorimotor information flow is flexible, state-dependent, and multidirectional. How does the brain integrate multiple sensory streams with the outcomes of past behavior to holistically guide action selection? How do memory and motivation shape sensorimotor processing, and how are they shaped by it? Inputs and outputs both shape representation in neural networks, so an integrated view is also needed to decipher the logic of sensorimotor coding. For example, to what extent does the animal's behavioral repertoire shape the priorities of sensory computation? These questions are becoming increasingly experimentally accessible [1, 2], yet they are formidable, and their answers will be complicated. We argue here that quantitative testable theories will be needed to organize empirical findings and prioritize future experiments [3].

One can conceptualize a successful theoretical science as a well-lit terrain (Figure 1a). Like someone metaphorically looking for lost keys only under a streetlamp, scientists are limited by the conceptual frameworks they use to understand the world. Scientists need

streetlamps to ensure that they not only see but also comprehend whatever is in front of them. We believe that theoretical principles provide the best streetlamps, as once found they can illuminate many phenomena quantitatively. We consider their light as concrete models and derivable results, and the illuminated terrain as the principle's domain of applicability. A famous example from the history of physics is the principle that mass can be neither created nor destroyed. Physicists extracted light from this principle by deriving and using the continuity equation, which quantifies how the amount of matter in a given region changes via inward and outward flow. The continuity equation in turn implied the diffusion equation, revealing this principle's wide domain of applicability to phenomena spanning statistical physics, chemistry, and biology.

Useful principles can have limited domains of applicability for two reasons. First, useful principles can fundamentally fail, as when Einstein realized that mass could be converted to energy. This particular failure led to generalization, as the principle of conservation of energy gracefully absorbed the conservation of mass. The generalized continuity equation could then be applied to new domains like cosmology. Second, a principle's domain of applicability is limited by its relevance to specific questions of interest. For example, the continuity equation contributes centrally to the theory of fluid flow but not rigid-body motion, even though it's valid in both settings. Moreover, P. W. Anderson argued in "More is Different" that understanding a complex system requires fundamentally different concepts and laws at each scale [4]. The multiscale nature of the brain is well appreciated [5], and we will probably need many streetlamps to illuminate sensorimotor processing. For instance, the conservation of mass can help a neuroscientist calculate the time required for neurotransmitter diffusion, but it predicts little about neuronal population coding. We must develop a good understanding of when each principle works, when it breaks down, and when it becomes irrelevant. Only then can we choose the right principles to analyze the question at hand.

In this review we highlight five theoretical principles that are shedding light onto brain-wide sensorimotor processing. In the next section we briefly illustrate the streetlamp analogy with two example principles that show the utility of familiar principles for brain-wide neuroscience. In the remaining three sections we discuss three more principles that we consider especially applicable to brain-wide recordings of sensorimotor processes.

Models of brain-wide phenomena can further reveal a principle's domain of applicability

We begin by illustrating the analogy with two familiar principles that have successfully driven experimental and theoretical progress in brain-wide neuroscience. First, much recent work has emphasized the principle that *neurons in motor regions serve primarily as participants in a large dynamical system that collectively generates bodily movements through ongoing neuronal dynamics* [6, 7] (Figure 1a). Single neuron tuning properties may or may not be interpretable, so it can be critical to characterize the dynamical system holistically. Kato et al. revealed new domains of applicability for this principle by using it to explain how the *C. elegans* brain generates stereotyped movement sequences [8] (Figure

1b). In particular, they attributed the activity of most neurons to shared low-dimensional dynamics that followed highly stereotyped periodic trajectories (Figure 1c). Different parts of these trajectories reliably mapped to different movements in the behavioral sequence (Figure 1d), and subtle differences between nearby trajectories corresponded to behavioral variations. Spontaneous and stimulus-evoked population dynamics were robust to single neuron silencing, highlighting globally distributed sensorimotor processing. This example principle thus allowed Kato et al. to illuminate brain-wide mechanisms of sequential behavior that were not localized to single neurons or brain areas.

A second familiar principle is that *cortical circuits operate in a state of balanced excitation and inhibition* [9]. The theoretical appeal of balanced excitation and inhibition is multifaceted and deep [10, 11], but the basic principle makes intuitive sense. Too much excitation destabilizes network dynamics, and too much inhibition prevents information transmission. Many computational properties of local cortical processing emerge in inhibition-stabilized networks [12, 13], a model variant of balance where strong excitation would lead to instability without inhibition. Joglekar et al. [14] cast light from the balance principle onto multi-regional cortical networks by generalizing the balanced amplification model of Murphy et al. [12] to a global form that balanced strong long-range excitation with local inhibition. Joglekar et al. found that their global balanced amplification model enhanced signal propagation throughout the cortical hierarchy [14], so balance might facilitate hierarchical sensorimotor processing in cortex.

Certain principles may be particularly useful for understanding brain-wide recordings of sensorimotor processing. We therefore devote each of the next three sections to highlighting a principle that has repeatedly proven its worth. Measurements of cellular-resolution brain-wide activity excite us because they can act as the keystone of a bridge spanning local circuits and systems. The next two sections explain this bridge through its top-down arc from larger-scale task variables to cellular activity and its bottom-up arc from smaller-scale biological mechanisms to the same activity. The final section discusses comparisons of brains to performance-optimized artificial neural networks.

Parsing brain-wide activity through quantitative algorithms of behavior

The empirical basis of the top-down arc from task variables to cellular activity is brain activity mapping. Single neuron activity is heterogeneously shaped by many environmental, behavioral, and cognitive task variables [8, 15–18], but a global organization coexists with this heterogeneity. A systems-scale understanding of the brain requires us to resolve how brain activity is simultaneously distributed across the brain and specialized within its subdivisions. The principle that *cellular activity, somewhere in the brain, represents task-variables linearly* underlies most brain activity mapping studies [9, 19]. In order to maximize the utility of this principle, researchers must build quantitative algorithmic models that explain how a multitude of behavioral, environmental, and cognitive variables unfold throughout the task. An important caveat is that correlative maps are not causal, and this principle does not imply that the task variables capture what matters most to the neural circuit dynamics and behavior.

Multiple recent papers focused on different aspects of the behavioral algorithm for the zebrafish optomotor response and thereby discovered a diversity of cellular participants in the sensorimotor transformation [20–25]. Here we illustrate the principle through Dragomir et al. [23] and Bahl et al. [24], who found congruent results by independently performing similar experiments and analyses. These authors recorded behavioral and neuronal responses to random dot stimuli with variable right-left motion coherence (Figure 2a). Momentary turning rates followed a continuous evidence accumulation model (Figure 2b), which temporally integrated stimuli over the previous few trials to explain long timescale behavior correlations (Figures 2c). The behavioral model thereby identified evidence accumulation as a covert computation that must occur in the brain (Figure 2b), leading the authors to map brain activity by the timescales on which neurons integrate directional motion signals (Figure 2d). This revealed significant neuronal diversity within and across brain regions, with some neurons responding rapidly to stimuli and others integrating signals within and across trials. Bahl et al. additionally found that the behavior was better explained if the behavior model included a dynamic decision threshold, and both sets of authors ultimately mapped their behavioral models onto neuronal signals by combining these and other results (Figure 2e).

Evidence accumulation is a canonical computation for decision making, and these zebrafish studies illustrate the synergy between behavioral modeling and cellular-resolution brain-wide recordings. Similar approaches will also be useful for dissecting sensorimotor decision-making processes in higher vertebrates. For example, several recent studies used sophisticated behavioral models to provide a nuanced view of rodent evidence accumulation [26, 27]. The authors' inferred behavioral algorithms permitted them to correlate neuronal activity in specific cortical regions with subtle cognitive variables [28], and brain activity mapping showed that decision-making variables are widely distributed [29]. As large-scale recording techniques continue their rapid rise [16, 17, 29, 30], behavior-model driven brain mapping will surely produce countless discoveries. Nevertheless, the domain of applicability of this principle is likely limited, as the brain may not linearly represent all relevant task variables [31, 32]. Conceptual and technical progress is needed to derive principled methodology for nonlinear brain activity mapping.

Making predictions with mechanistic models: finding constraints amongst uncertainty

Mapping sensory, cognitive, and behavioral variables does not explain how the brain actually builds these representations from related sensory inputs and motor outputs. This is the purview of the bottom-up arc from biological mechanism to cellular-resolution brain-wide activity. One generally hopes to construct a mechanistically detailed dynamical system model that can elucidate the causal flow of information throughout the sensorimotor network, and several recent studies followed brain-wide activity mapping with multi-regional neural network modeling [20, 24, 33, 34]. Complete system specification is challenging [35, 36], but the principle that *relatively few combined parameters control emergent properties in high-dimensional physical systems* [3, 37] provides a guide for nevertheless designing, constraining, and testing large-scale neural network models [38].

This principle has two facets relevant for mechanistically understanding distributed neural networks. First, many cellular and circuit mechanisms may be irrelevant for explaining the neuronal responses underlying specific sensorimotor transformations [20, 36, 39]. For instance, brain-wide sensorimotor circuits overlap with each other and operate concurrently with unrelated behavioral and cognitive processes [8, 15, 16, 40–42]. Some low-level mechanisms will only be needed to generate responses in these unrelated contexts. Models can therefore be simplified by ignoring these irrelevant details. Second, other mechanistic details may be critical for explaining the functional responses in focus [43]. This means that incompletely constrained models may still be able to make accurate experimental predictions.

Many neural network models are parametrized by their synaptic weights. Critical synaptic connections could provide compelling experimental predictions, but discerning critical model parameters is generally hard computationally [43–45]. This further increases the appeal of simple models [3], especially when analytical methods can relate synaptic weights and neuronal responses comprehensively [46]. Biswas and Fitzgerald [47] recently developed a theoretical formalism to find synaptic connections required for stimulus responses in threshold-linear recurrent neural networks (Figure 3a), a rich nonlinear model class with promising analytical tractability [48]. The authors first analytically found all synaptic weight matrices with specified steady state neural responses (Figure 3b). They then sought critical synaptic connections that must be present (Figures 3c). When compared to linear neural networks, threshold nonlinearities increased the size of the solution space by permitting variability in “semi-constrained” dimensions that appear when a neuron does not respond to a stimulus (Figure 3d). Nevertheless, the authors derived a geometric formula that could analytically pinpoint many critical synaptic connections, even in large networks. This theoretical approach could thus be used to make rigorous anatomical predictions from brain-wide neuronal responses [1].

Performance-optimized artificial neural networks as models of brain-wide computation

We now turn to the fifth and final principle. Analyzing brain-wide computation is complicated by our limited knowledge of the brain’s high-level functions and constituent mechanisms, with mechanistic access most limited for large animals. However, the principle that *evolutionary convergence of neural computations highlights computational mechanisms suited to solve general problems under common constraints* allows neuroscientists to understand complex brains via comparisons to simpler ones. An interesting twist of this principle treats artificial neural networks (ANNs) as uniquely tractable model systems, whose functions and mechanisms are fully known [49, 50]. ANNs currently achieve remarkable performance in multiple artificial intelligence tasks. Such tasks require ANNs to transform low-level sensory information into task-relevant representations that support goal-directed outputs. Biological neural networks (BNNs) perform analogous tasks using neurons that span much of the brain, so comparing network-wide sensorimotor processing in ANNs and BNNs may provide insights into many brain areas [51, 52]. Not all aspects

of ANNs will translate to BNNs, but the same is true when comparing two BNNs from different animal species [25, 53, 54].

Recent work by Haesemeyer et al. shows the utility of this comparative approach [51]. The authors trained an ANN to predict temperature changes enabling a virtual agent to navigate heat gradients. The authors then compared activity in the ANN to activity previously recorded in larval zebrafish [33, 51]. Neuron types emerged in the ANN with response properties matched to the BNN [51]. This permitted experiments in the ANN that made predictions for the BNN. For example, targeted ablation of some ANN neuron types substantially reduced the agent's navigational abilities, and the authors predicted that these neuron types are critical in the BNN. Most impressively, the ANN predicted a critical neuron type that was not previously seen in the zebrafish brain. This prediction prompted the authors to reanalyze their brain-wide imaging data and discover a neuron type in the cerebellum that matched the ANN. This illustrates why brain-wide neuronal recordings are valuable for facilitating comparisons to large-scale ANNs. The authors did not know *a priori* where to look.

The notion that representations converge in ANNs and BNNs is increasingly familiar, as neuroscientists have repeatedly found this outcome in sensory and cognitive neuroscience [55–57]. If neuroscientists can understand how ANN representation emerges from its optimization function, network architecture, and learning dynamics [50], then ANN modeling will become an immensely useful tool for examining the principles of comparative computational neuroscience. It is therefore crucial to train ANNs using many learning paradigms, optimization criteria, and regularization schemes. For example, studies have ascribed importance to regularization that promotes “simple” activity patterns in ANNs [58]. Convergences could be enhanced by the development of ANNs that explicitly respect biological findings [59]. The field is probably only beginning to glimpse the full potential of this comparative approach.

Outlook

In this review, we've likened the importance of theoretical principles to that of streetlamps. Scientists can get lost in the dark without them, and beautiful things may go unnoticed if hidden in shadow. We've provided examples of five principles that have utility in brain-wide neuroscience, and each principle had a broad but limited domain of applicability. Neuroscientists will thus have to bring these and other principles together to understand brain-wide sensorimotor circuits. We would refer to the resulting theoretical system as a framework. Frameworks can be far more illuminating than any of their constituent principles. For example, Galileo's law of inertia was an insightful principle that changed how his contemporaries saw motion, but it became inestimably more powerful once absorbed into the framework of Newtonian mechanics as its first law. In neuroscience, the Hodgkin-Huxley framework is exceptionally powerful because it coherently integrated many principles of neuronal excitability whose concepts fundamentally changed how experimenters interpreted their data. We are optimistic that the principles highlighted here are synergistic and complementary. They may therefore contribute productively to whatever

theoretical framework eventually systematizes the fundamentals of brain-wide sensorimotor computation.

Acknowledgments:

The authors thank Jing-Xuan Lim and Diana Snyder for posing for Fig. 1a and Misha Ahrens, Ann Hermundstad, Eva Naumann, and Andrew Saxe for discussion and comments on the manuscript. This work was supported by the Howard Hughes Medical Institute.

References

1. Koroshetz W, et al. , The State of the NIH BRAIN Initiative. *Journal of Neuroscience*, 2018. 38(29): p. 6427–6438. [PubMed: 29921715]
2. Calarco JA and Samuel ADT, Imaging whole nervous systems: insights into behavior from worms to fish. *Nature Methods*, 2019. 16(1): p. 14–15. [PubMed: 30573822]
3. Bialek W, Perspectives on theory at the interface of physics and biology. *Reports on Progress in Physics*, 2018. 81(1).
4. Anderson PW, More Is Different - Broken Symmetry and Nature of Hierarchical Structure of Science. *Science*, 1972. 177(4047): p. 393-&. [PubMed: 17796623]
5. Churchland PS and Sejnowski TJ, Perspectives on cognitive neuroscience. *Science*, 1988. 242(4879): p. 741–5. [PubMed: 3055294]
6. Shenoy KV, Sahani M, and Churchland MM, Cortical control of arm movements: a dynamical systems perspective. *Annu Rev Neurosci*, 2013. 36: p. 337–59. [PubMed: 23725001]
7. Sauerbrei BA, et al. , Cortical pattern generation during dexterous movement is input-driven. *Nature*, 2020. 577(7790): p. 386–391. [PubMed: 31875851] • This study applied the advanced tools of rodent neuroscience to assess whether motor cortex was an autonomous or input-driven dynamical system.
8. Kato S, et al. , Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell*, 2015. 163(3): p. 656–69. [PubMed: 26478179]
9. Shadlen MN and Newsome WT, Noise, neural codes and cortical organization. *Curr Opin Neurobiol*, 1994. 4(4): p. 569–79. [PubMed: 7812147]
10. van Vreeswijk C and Sompolinsky H, Chaotic balanced state in a model of cortical circuits. *Neural Comput*, 1998. 10(6): p. 1321–71. [PubMed: 9698348]
11. Deneve S and Machens CK, Efficient codes and balanced networks. *Nat Neurosci*, 2016. 19(3): p. 375–82. [PubMed: 26906504]
12. Murphy BK and Miller KD, Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 2009. 61(4): p. 635–48. [PubMed: 19249282]
13. Rubin DB, Van Hooser SD, and Miller KD, The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 2015. 85(2): p. 402–17. [PubMed: 25611511]
14. Joglekar MR, et al. , Inter-areal Balanced Amplification Enhances Signal Propagation in a Large-Scale Circuit Model of the Primate Cortex. *Neuron*, 2018. 98(1): p. 222–234 e8. [PubMed: 29576389] •• This study used novel anatomical data on the strength of excitatory pathways in the primate brain to model how excitation and inhibition could balance at the brain scale. Their model revealed benefits of balanced amplification for inter-areal signal propagation that will likely be relevant for understanding cortical networks for sensorimotor processing.
15. Chen XY, et al. , Brain-wide Organization of Neuronal Activity and Convergent Sensorimotor Transformations in Larval Zebrafish. *Neuron*, 2018. 100(4): p. 876–+. [PubMed: 30473013]
16. Stringer C, et al. , Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 2019. 364(6437): p. 255–+. [PubMed: 31000656]
17. Steinmetz NA, et al. , Distributed coding of choice, action and engagement across the mouse brain. *Nature*, 2019. 576(7786): p. 266–+. [PubMed: 31776518]

18. Randlett O, et al. , Whole-brain activity mapping onto a zebrafish brain atlas. *Nature Methods*, 2015. 12(11): p. 1039–1046. [PubMed: 26778924]
19. DiCarlo JJ and Cox DD, Untangling invariant object recognition. *Trends in Cognitive Sciences*, 2007. 11(8): p. 333–341. [PubMed: 17631409]
20. Naumann EA, et al. , From Whole-Brain Data to Functional Circuit Models: The Zebrafish Optomotor Response. *Cell*, 2016. 167(4): p. 947–+. [PubMed: 27814522]
21. Kawashima T, et al. , The Serotonergic System Tracks the Outcomes of Actions to Mediate Short-Term Motor Learning. *Cell*, 2016. 167(4): p. 933–+. [PubMed: 27881303]
22. Mu Y, et al. , Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior. *Cell*, 2019. 178(1): p. 27–+. [PubMed: 31230713] • This study discovered a beautiful role for glial cells in inducing a brain-wide transition in neuronal activity that accompanies the onset of behavioral passivity.
23. Dragomir EI, Stih V, and Portugues R, Evidence accumulation during a sensorimotor decision task revealed by whole-brain imaging. *Nature Neuroscience*, 2020. 23(1): p. 85–+. [PubMed: 31792463] •• Back-to-back with Bahl et al., this paper showed that larval zebrafish accumulate evidence to make behavioral decisions in uncertain environments. This paper makes great use of cellular-resolution brain activity maps.
24. Bahl A and Engert F, Neural circuits for evidence accumulation and decision making in larval zebrafish. *Nature Neuroscience*, 2020. 23(1): p. 94–+. [PubMed: 31792464] •• Back-to-back with Dragomir et al., this paper showed that larval zebrafish accumulate evidence to make behavioral decisions in uncertain environments. This paper makes great use of behavioral model comparisons.
25. Yildizoglu Y, et al. , A neural representation of naturalistic motion-guided behavior in the zebrafish brain. *Current Biology*, 2020. 30(12): p. 2321–2333.e6. [PubMed: 32386533]
26. Brunton BW, Botvinick MM, and Brody CD, Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science*, 2013. 340(6128): p. 95–98. [PubMed: 23559254]
27. Pinto L, et al. , An Accumulation-of-Evidence Task Using Visual Pulses for Mice Navigating in Virtual Reality. *Frontiers in Behavioral Neuroscience*, 2018. 12.
28. Hanks TD, et al. , Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, 2015. 520(7546): p. 220–U195. [PubMed: 25600270]
29. Koay SA, et al. , Neural Correlates of Cognition in Primary Visual versus Neighboring Posterior Cortices during Visual Evidence-Accumulation-based Navigation. *bioRxiv*, 2019: p. 568766. • This work builds on a previous behavioral model to map the cortical correlates of visual evidence accumulation in mice.
30. Sofroniew NJ, et al. , A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife*, 2016. 5.
31. Shamir M and Sompolinsky H, Nonlinear population codes. *Neural Computation*, 2004. 16(6): p. 1105–1136. [PubMed: 15130244]
32. Schneidman E, Towards the design principles of neural population codes. *Current Opinion in Neurobiology*, 2016. 37: p. 133–140. [PubMed: 27016639]
33. Haesemeyer M, et al. , A Brain-wide Circuit Model of Heat-Evoked Swimming Behavior in Larval Zebrafish. *Neuron*, 2018. 98(4): p. 817–831 e6. [PubMed: 29731253]
34. Andalman A, et al. , Neuronal Dynamics Regulating Brain and Behavioral State Transitions. *Cell*, 2019. 177(4): p. 970–+. [PubMed: 31031000]
35. Markram H, et al. , Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 2015. 163(2): p. 456–92. [PubMed: 26451489]
36. Billeh YN, et al. , Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex. *Neuron*, 2020. • This paper models visual cortex with biophysically detailed neurons and integrate-and-fire neurons, showing that both model networks have similar dynamics.
37. Transtrum MK, et al. , Perspective: Slowness and emergent theories in physics, biology, and beyond. *J Chem Phys*, 2015. 143(1): p. 010901. [PubMed: 26156455]
38. O’Leary T, Sutton AC, and Marder E, Computational models in the age of large datasets. *Curr Opin Neurobiol*, 2015. 32: p. 87–94. [PubMed: 25637959]

39. Fisher D, et al. , A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. *Neuron*, 2013. 79(5): p. 987–1000. [PubMed: 24012010]
40. Robie AA, et al. , Mapping the Neural Substrates of Behavior. *Cell*, 2017. 170(2): p. 393–406 e28. [PubMed: 28709004]
41. Aimon S, et al. , Fast near-whole-brain imaging in adult *Drosophila* during responses to stimuli and behavior. *PLoS Biol*, 2019. 17(2): p. e2006732. [PubMed: 30768592]
42. Mann K, Gallen CL, and Clandinin TR, Whole-Brain Calcium Imaging Reveals an Intrinsic Functional Network in *Drosophila*. *Curr Biol*, 2017. 27(15): p. 2389–2396 e4. [PubMed: 28756955]
43. Marder E and Taylor AL, Multiple models to capture the variability in biological neurons and networks. *Nat Neurosci*, 2011. 14(2): p. 133–8. [PubMed: 21270780]
44. Bittner SR, et al. , Interrogating theoretical models of neural computation with deep inference. *bioRxiv*, 2019: p. 837567.
45. Gonçalves PJ, et al. , Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, 2020. 9: p. e56261. [PubMed: 32940606]
46. Baldi P and Hornik K, Neural Networks and Principal Component Analysis - Learning from Examples without Local Minima. *Neural Networks*, 1989. 2(1): p. 53–58.
47. Biswas T and Fitzgerald JE, A geometric framework to predict structure from function in neural networks. *arXiv*, 2020. •• This theoretical paper uses an ensemble modeling approach to determine anatomical structures needed to generate functional responses in threshold-linear recurrent neural networks. It finds that a relatively small number of neuronal response patterns can suffice for making precise and testable anatomical predictions.
48. Morrison K, et al. , Diversity of emergent dynamics in competitive threshold-linear networks: a preliminary report. *arXiv*, 2016.
49. Rogers TT and McClelland JL, Parallel Distributed Processing at 25: further explorations in the microstructure of cognition. *Cogn Sci*, 2014. 38(6): p. 1024–77. [PubMed: 25087578]
50. Richards BA, et al. , A deep learning framework for neuroscience. *Nature Neuroscience*, 2019. 22(11): p. 1761–1770. [PubMed: 31659335]
51. Haesemeyer M, Schier AF, and Engert F, Convergent Temperature Representations in Artificial and Biological Neural Networks. *Neuron*, 2019. 103(6): p. 1123–+. [PubMed: 31376984] •• This paper compares neural activity in the larval zebrafish brain to activity in an artificial neural network trained to navigate temperature gradients. The artificial neural network predicted an unexpected neuronal response type that was experimentally verified by reanalysis of published brain-wide neuronal recordings.
52. Merel J, et al. , Deep neuroethology of a virtual rodent, in *ICLR 2020*. 2020. • This study uses advanced methods from deep reinforcement learning to train a virtual rodent to accurately control its motor behavior in ethologically relevant tasks.
53. Clark DA, et al. , Flies and humans share a motion estimation strategy that exploits natural scene statistics. *Nature Neuroscience*, 2014. 17(2): p. 296–303. [PubMed: 24390225]
54. Nitzany EI, et al. , Neural computations combine low- and high-order motion cues similarly, in dragonfly and monkey. *bioRxiv*, 2017: p. 240101.
55. Yamins DLK and DiCarlo JJ, Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 2016. 19(3): p. 356–365. [PubMed: 26906502]
56. Yang GR, et al. , Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci*, 2019. 22(2): p. 297–306. [PubMed: 30643294]
57. Whittington JC, et al. , The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. *bioRxiv*, 2019: p. 770495.
58. Sussillo D, et al. , A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience*, 2015. 18(7): p. 1025–+. [PubMed: 26075643]
59. Litwin-Kumar A and Turaga SC, Constraining computational models using electron microscopy wiring diagrams. *Current Opinion in Neurobiology*, 2019. 58: p. 94–100. [PubMed: 31470252]

HIGHLIGHTS:

- Multiple theoretical principles are needed to understand brain-wide neural activity
- Brain-wide recordings provide new opportunities for testing familiar principles
- Quantitative behavioral models reveal target variables for brain activity mapping
- Sensitivity analyses aid mechanistic predictions from highly parameterized models
- Artificial neural networks can serve as tractable model systems for neuroscience

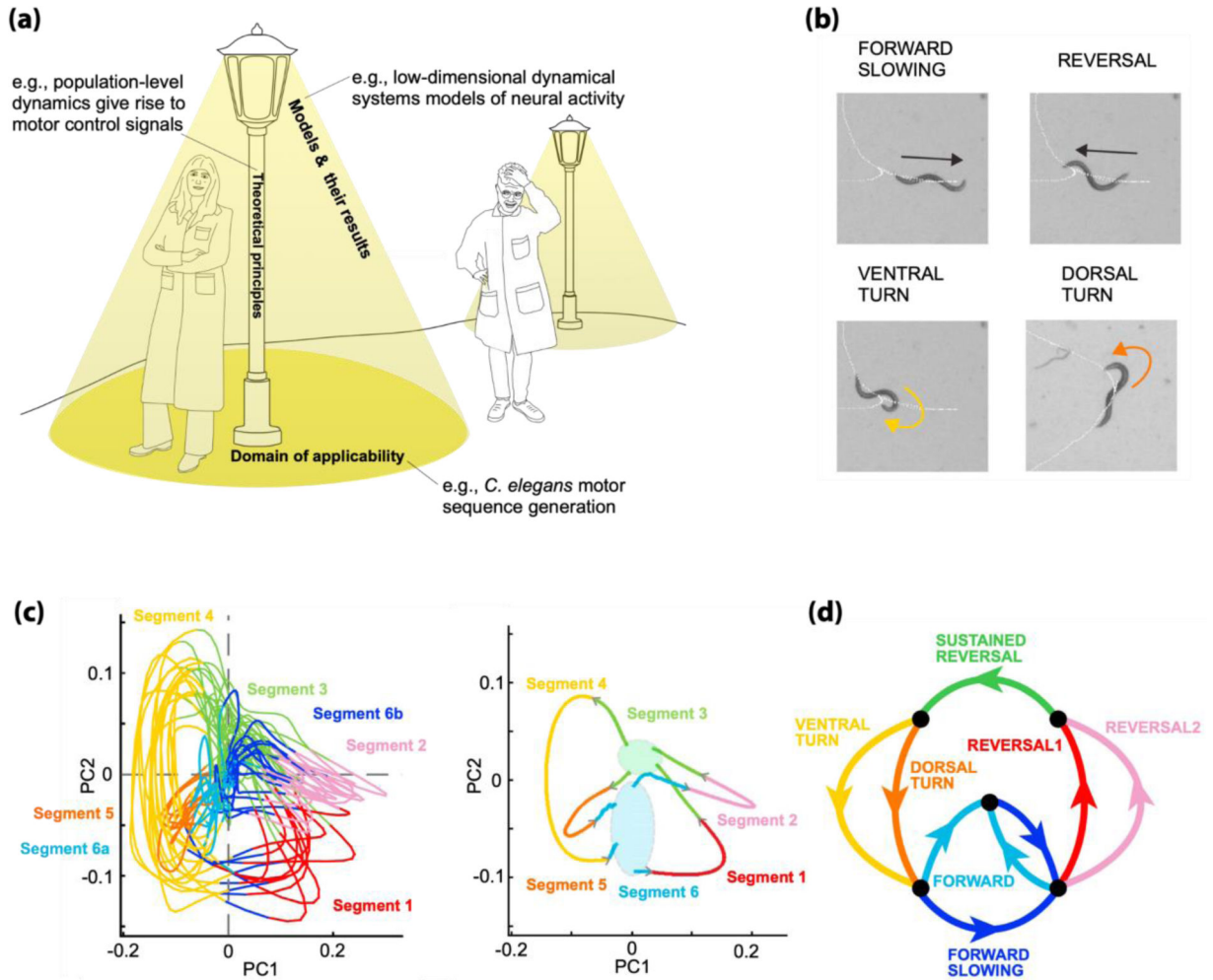


Figure 1:
(a) We conceptualize theoretical principles as streetlamps, related models and their derivable results as emitted light, and the empirical domain of applicability as the illuminated terrain. For example, the principle that neurons involved in movement production serve primarily to support the ongoing dynamics of a large dynamical system has led to low-dimensional dynamical systems models that provide means of interpreting the activity of large populations of neurons. This principle was articulated in the domain of primate motor cortex and recently applied by Kato et al. [8] to study the control of motor-sequences in *C. elegans*. Multiple streetlamps are required to illuminate the full landscape of neuroscience phenomena. The job of the neuroscientist is hardest when they find themselves outside the guiding light of all available streetlamps. **(b)** Kato et al. recorded brain-wide neuronal activity from immobilized *C. elegans* fictively performing stereotyped movement sequences known as pirouettes. A pirouette consists of a reversal from forward to backward crawling, followed by a turn along either the dorsal or ventral body axis, and finally by the resumption of forward crawling. **(c, left)** Principal component analysis (PCA) on the brain-wide recordings revealed trajectories in neural population activity space that cyclically followed stereotyped paths. By analyzing single neuron and population activity, the authors were able to decompose neural state space trajectories into several segment types (colors). See

Kato et al. for details. (c, right) Average segment trajectories illustrate the brain's cyclical dynamics. Ovals indicate regions where different bundles of the trajectories mix. They serve to graphically connect segment averages to each other. (d) The individual segment types of the identified brain cycle could be reliably mapped to the major components of the worm's movement sequence (colors coded to match (c)). This correspondence was made by comparing brain-wide neuronal recordings during fictive behavior to single neuron recordings in freely moving worms.

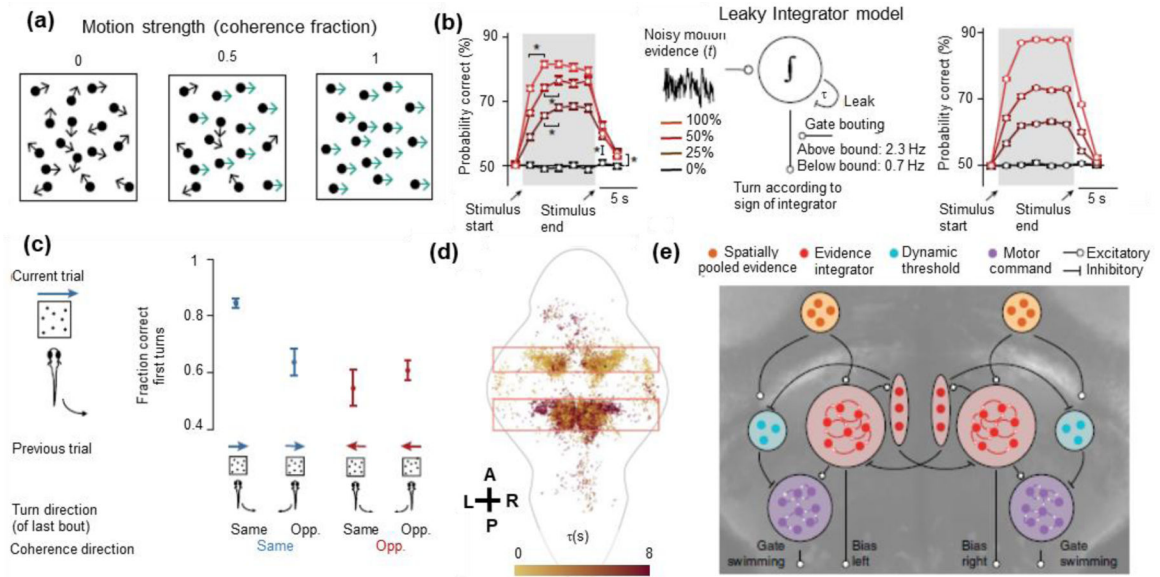
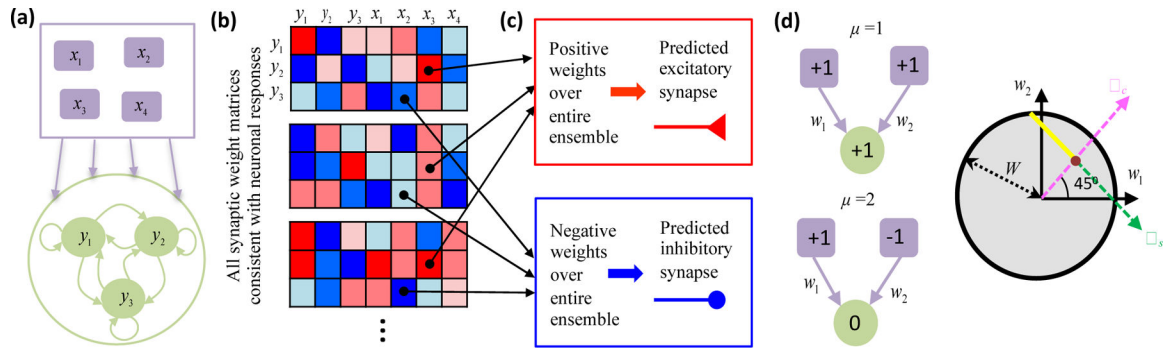


Figure 2:

Mapping distributed sensorimotor processing in larval zebrafish using an evidence accumulation model. **(a)** Dragomir et al. [23] and Bahl et al. [24] used visual motion stimuli to drive turning behavior. Stimuli with three different strengths of rightward motion are shown (coherence fraction = 0, 0.5, 1). Panel from [23]. **(b, left)** The probability of correct turns (*i.e.* turns in the direction of motion) increased over time within the trial and also increased with higher coherence fractions. **(b, middle)** These results suggest an evidence accumulation model, whereby zebrafish leakily integrate noisy motion evidence to determine the direction of visual motion and set the rates of leftward and rightward turning. **(b, right)** The model successfully reproduces the observed behavioral patterns. Panel from [24]. **(c)** Motor actions do not reset the evidence accumulator, and its integration time scale is long enough to span several bouts and trials. The fraction of correct first turns accordingly depended on the direction of visual motion in the preceding trial and on the direction of the last turn. Panel from [23]. **(d)** Implementing the leaky integrator model would require the brain to first detect fast changes in sensory signals and then integrate them temporally. The authors therefore mapped brain activity by the integration time constant that best modeled each neuron's stimulus responses. The map of model time constants showed a diversity of neuronal responses that were organized anatomically and could realize the computational steps of the behavioral models. Each colored dot is a segmented neuron, A-P denotes the anterior-posterior axis, and L-R denotes the left-right axis. Also shown are an outline of the zebrafish brain and boxes that highlight regions with many task-relevant neurons. Panel from [23]. **(e)** The authors ultimately superimposed their behavioral models onto cellular-resolution brain-wide maps, revealing neurons that detected motion signals in each direction, neurons that competitively pooled leftward and rightward motion signals, neurons that dynamically set the threshold separating low and high values of accumulated evidence, and neurons that represented stochastically generated leftward and rightward motor turning commands. Panel from [24].

**Figure 3:**

(a) Biswas and Fitzgerald [47] first specified steady-state responses of a recurrent threshold-linear neural network (green neurons) receiving feedforward input (purple neurons). (b) Assuming that the number of specified responses does not exceed the number of synapses onto each neuron, they then found all synaptic weight matrices having fixed points at the specified responses. Red (blue) matrix elements denote positive (negative) synaptic weights. (c) When a weight was consistently positive (or consistently negative) across all possibilities, then the model needed a nonzero synaptic connection to generate the responses. The model thus predicts that this synapse should exist experimentally. (d) The threshold nonlinearity provides flexibility, but some features of connectivity can nevertheless be definite. (d, left) Cartoon depicting steady state response patterns of a simple feedforward network under two stimulus conditions. The driven neuron responds in one stimulus condition (top) but is silent in the other (bottom). Numbers inside the neurons indicate activity levels. (d, right) The circle represents the two-dimensional space of allowed synaptic weights, assuming that the total weight norm is bounded by W . The two response conditions would uniquely determine the weight magnitudes in a linear neural network (brown dot, $w_1 = w_2 = 1/2$). However, the driven neuron's silent response in the threshold-linear neural network only implies that the synaptic weight component along the ω_s direction is negative, because any sub-threshold input would suffice. All points on the semi-infinite yellow line thus reproduce the observed responses. Biswas and Fitzgerald refer to the ω_s direction as a "semi-constrained" dimension in the solution space. Despite this flexibility, all model solutions require $w_2 > 0$.