Harnessing behavioral diversity to understand neural computations for cognition

Simon Musall1, Anne E Urai1, David Sussillo2,3,4 and Anne K Churchland1

With the increasing acquisition of large-scale neural recordings comes the challenge of inferring the computations they perform and understanding how these give rise to behavior. Here, we review emerging conceptual and technological advances that begin to address this challenge, garnering insights from both biological and artificial neural networks. We argue that neural data should be recorded during rich behavioral tasks, to model cognitive processes and estimate latent behavioral variables. Careful quantification of animal movements can also provide a more complete picture of how movements shape neural dynamics and reflect changes in brain state, such as arousal or stress. Artificial neural networks (ANNs) could serve as artificial model organisms to connect neural dynamics and rich behavioral data. ANNs have already begun to reveal how a wide range of different behaviors can be implemented, generating hypotheses about how observed neural activity might drive behavior and explaining diversity in behavioral strategies.

Addresses
1 Cold Spring Harbor Laboratory, Neuroscience, Cold Spring Harbor, NY, USA
2 Google AI, Google, Inc., Mountain View, CA, USA
3 Department of Electrical Engineering, Stanford University, Stanford, CA, USA
4 Stanford Neurosciences Institute, Stanford University, Stanford, CA, USA

Corresponding author: Churchland, Anne K (churchland@cshl.edu)
1 They are co-first authors and contributed equally to this work.

Current Opinion in Neurobiology 2019, 58:229–238
This review comes from a themed issue on Computational neuroscience
Edited by Máté Lengyel and Brent Doiron

0959-4388/© 2019 Elsevier Ltd. All rights reserved.

Introduction
To understand the computations implemented by neural network dynamics, it is crucial to study them in the context of the behavioral output they generate. For example, studies of decision-making usually leverage behavioral tasks in which animals are trained to produce a specific behavioral response following presentation of sensory stimuli. This allows repeated measurements of neural activity with systematic manipulation of the inputs, supporting the ability to map the neural pathways that transform sensory inputs into action [1–4]. However, tasks that solely rely on simple behaviors, such as licking in response to a sensory stimulus, limit the diversity of the observable neural dynamics (Box 1), making it difficult to estimate whether they accurately represent neural function under more complex conditions [5–7]. For example, focusing only on binary choices might obscure ongoing behaviors or the animal’s brain state, both of which strongly affect neural activity and task performance [8,9,10,11,12,13*].

Recent experimental, analytical and theoretical advances provide opportunities to overcome these issues, bolstering our ability to connect neural activity to behavior. Here, we provide an overview of these emerging methods. We argue that the decision-making field should embrace behavioral complexity as a way to understand fluctuations in neural activity, gain insight into an animal’s brain state and distinguish behavioral strategies. First, we describe task features that increase behavioral complexity and allow experimenters to infer the animal’s estimate of computationally relevant latent behavioral variables. The latter are not directly measured but can be derived from behavioral models. Second, we highlight new ways of quantifying animal movements and behavioral motifs, and describe how such data can aid the interpretation of rich behaviors and individual task strategy (Figure 1), as well as single-trial neural data. Lastly, we highlight the use of artificial neural networks (ANNs), especially multi-layered networks known as deep neural networks, as a way to create hypotheses for how high-dimensional neural dynamics can give rise to behavior. ANNs complement traditional latent variable models that arise from behavioral modeling. They can be viewed as simple artificial model organisms for which the entire connectome and activations are observable, and for which many thousands of networks can be trained and studied. ANNs are also starting to provide insight into individual differences in behavioral strategy, even for agents that pursue the same experimental goal (Figure 1) [14*].

Task features to estimate latent behavioral variables
In the study of decision-making, animals often perform a behavioral task designed to engage a cognitive process of
interest. This task-based approach can support the extraction of latent behavioral variables (Table 1), which are not directly observable from behavior but inferred through mathematical models of cognitive processes. Such models are a workhorse of mathematical psychology and cognitive computational neuroscience [15], and can capture an agent’s observed behavior so that its components can be related to neural activity (Figure 2). They encompass both ideal-observer models and descriptive process models, such as sequential sampling [16,17], reinforcement learning [18,19] and Bayesian ideal observer models [20,21]. Latent behavioral variables extracted from these models, such as the inferred accumulation of sensory evidence [17], have been successfully linked to brain function at different levels, from single cells [23,24], microcircuits [25,26], to macro-scale neural mass signals [27,28] and high-dimensional state representations of neural population activity [29].

Importantly, understanding the neural basis of cognitive processes is possible only when behavior allows the experimenter to estimate computationally relevant latent variables. This estimation critically depends on the specific design of the behavioral task. But which task features should be incorporated to effectively estimate latent behavioral variables?

Figure 1

Animals can exhibit a diverse range of behaviors and strategies even when solving the same task. Insight into this diversity might come from increasing task complexity, detailed quantification of animal behavior and examination of ANNs trained to solve the same problem. Image by Julia Kuhl.
Stimulus features can be used to fit models that estimate latent behavioral variables. For instance, stochastic, time-varying stimuli allow the use of behavioral models to infer the animal’s time-varying estimate of accumulated evidence; these can then be related to neural activity [30,31]. Varying the duration of within-trial task episodes, such as a stimulus sequences, can further constrain parameter estimates, such as time constants of integration [32,33]. Presenting different sensory modalities at varying levels of reliability allows one to estimate their respective weight during multisensory integration [34] and evaluate how these weights are encoded in population neural activity [35].

Animal responses can likewise be used to estimate latent behavioral variables. For instance, in models of evidence accumulation, reaction times provide an estimate of the time for a decision variable to reach a bound [22,36] and post-choice waiting times provide an estimate of decision confidence [37]. Continuous responses such as reaching movements, moving a wheel or navigating in virtual reality can provide additional insight into the evolving

Table 1
Definitions of new terms emerging in the study of cognitive and behavioral circuits

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Examples</th>
<th>Biological versus artificial networks?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain state</td>
<td>Internally generated neural dynamics that fluctuate spontaneously, often in ways that are related to bodily constraints. These are often measured through physiological markers.</td>
<td>Arousal, fear, stress, hunger, motivation, engagement, drowsiness</td>
<td>Biological: [59]</td>
</tr>
<tr>
<td>Latent behavioral variable</td>
<td>The agent’s estimate of a computationally relevant quantity. These latent variables are inferred via behavioral models.</td>
<td>Accumulated evidence, bias, value, confidence</td>
<td>Both (biological: [22] ANN: [51])</td>
</tr>
<tr>
<td>Behavioral motif</td>
<td>A stereotyped series of movements that identifies a specific behavior. Ongoing behavior can be described as a continuous sequence, switching from one motif to the next.</td>
<td>Grooming, eating, walking, running, jumping, rearing, climbing, reaching</td>
<td>Both (biological: [60] ANN: [51])</td>
</tr>
<tr>
<td>Task-based approach</td>
<td>Experiments with an experimenter-defined task and a repeatable trial-structure. Non-human experiments usually include animals trained to perform an arbitrary movement to receive a reward.</td>
<td>Random dot motion task, reaching task, maze navigation task, image classification task</td>
<td>Both (biological: [34] ANN: [61**])</td>
</tr>
<tr>
<td>Observational approach</td>
<td>Experiments where self-generated behavior of untrained animals is observed and analyzed.</td>
<td>Home cage exploration, mating behavior, head-fixed wheel locomotion, place field mapping</td>
<td>Biological: [62] (currently)</td>
</tr>
<tr>
<td>Individual differences</td>
<td>Differences in an individual’s behavioral repertoire that allow to distinguish different animal types.</td>
<td>Exploratory versus fearful, dominant versus submissive, social versus asocial, active versus inactive</td>
<td>Both (biological: [63**] ANN: [14**])</td>
</tr>
</tbody>
</table>
decision process [29,38–40]. For example, changes in head orientation are related to upcoming animal choices and trial-to-trial neural variability [41]. Further, reaching trajectories in a virtual maze task can reveal changes of mind in free-choice trials [29].

As animals naturally integrate information across many trials to combine sensory information with prior knowledge and beliefs [42], variations in task history can reveal the algorithms by which agents integrate information across trials [43], learn the structure of their environment [44], perform probabilistic inference [45], and adjust their decision policy after errors [46]. For example, task features that span many trials, such as stimulus-response contingencies, reward probabilities or specific trial sequences, provide another opportunity to reveal latent behavioral variables. A model’s across-trial latent variables can then be linked to activity in specific brain regions [47] and single cells [48]. Dynamic logistic regression models also allow data-driven estimation of across-trial latent variables, such as an animal’s reliance on trial history or its choice bias, over long timescales and without assuming a specific generative model [49].

Rich stimuli, responses and task structure are crucial when studying the behavior of biological as well as artificial agents. Training animals and ANNs on the same task (Figure 2) can elucidate how computational strategies arise from certain task constraints, neural architectures and cost functions, and show how these computations are implemented using the elementary building blocks of neural networks. Knowledge of latent behavioral variables can be linked to internal ANN dynamics that underlie such task-relevant computations [4,50–55]. By including more detailed behavioral quantification and task complexity, the behavior of ANNs can be further constrained to generate network dynamics that are more comparable to biological neural circuits [56**]. ANNs may also complement traditional behavioral models to capture the complexity of richer or more ethological tasks [57].

Quantification of animal behavior

The earliest descriptions of animal behavior were generated by ethologists taking an observational approach (Table 1): they formulated a set of criteria (usually describing a sequence of simpler movements) to identify a specific behavioral motif (e.g. feeding or grooming;
Table 1) and quantify its occurrence [64]. While quantifying behavioral motifs used to be extremely laborious and susceptible to human error, technical advances have largely shifted the field towards automatic identification and quantification of pre-defined behavioral motifs [62]. With the rise of neural recordings in freely moving animals, such behavioral motifs are now routinely related to neural activity patterns [65,66].

There is an extensive toolbox for characterizing movements. This includes audio recording [67] and RFID tagging [68], and even more striking (and more applicable for decision-making), methods based on video data. New toolboxes utilize supervised deep learning methods (based on human-labeled examples) to automatically track the position and posture of animals as they navigate through their environment [69,70]. Given enough training data, these algorithms are remarkably robust, and can be readily applied to obtain readouts like movement velocity, spatial position, and body orientation from video data.

An extension of animal movement quantification is the unsupervised classification of behavioral motifs without a preconceived, observation-based template. Analyses have been optimized for C. elegans [71,72], larval zebrafish [73], and drosophila melanogaster [74,75], along with freely moving mice [60]. Here, time-frequency analysis or auto-regressive models can discover and quantify the occurrence of stereotyped temporal dynamics in low-dimensional movement representations and cluster them into distinct behavioral motifs. This unsupervised, data-driven approach can reveal diverse behavioral motifs, their frequency of occurrence and their sequential order [60,76], all of which can be linked to neural activity. Automated behavioral quantification has also begun to overturn long-held assumptions about the complexity of behavior. For instance, unsupervised classification methods applied to large databases of drosophila songs exposed song modes beyond the two that were long thought to make up the animal’s song repertoire [77].

While detailed movement quantification is now routinely applied in the observational approach, it has been argued that the behavior of well-trained animals might be overconstrained and consist largely of limited, non-ethological movements [7]. However, new analyses have uncovered that head-fixed mice explore a large array of uninstructed movements, such as whisking, locomotion and facial movements [12,13]. This shows that, even under highly constrained conditions (head-fixation, low-dimensional task contingencies), animal behavior is surprisingly rich and includes diverse uninstructed movements. While ethological validity and task complexity are important factors to consider when designing a behavioral task [7,8], it is intriguing that highly trained, head-fixed animals are far from being ‘task automats’. Instead, even as they become task experts, animals retain a richly varied array of instructed and uninstructed movements.

These uninstructed movements might seem at first to pose a major challenge for interpreting neural activity. Fortunately, they can be quantified using dimensionality reduction of video data. Despite not being required for task performance, uninstructed movements strongly modulate neural activity [8,9,10–12]. During a decision-making task, video-based movement representations are also closely related to neural population dynamics: they outweigh the importance of task-related variables like sensory stimuli or animal choice to predict neural activity on single trials [13]. Even in studies that do not explicitly focus on movements, movement quantifications can account for trial-by-trial fluctuations in neural activity that could otherwise be interpreted as spontaneous ‘noise’ [12,13] (Figure 2). Movement quantifications can also help interpret neural activity by uncovering neural response features, such as a sensitivity to sensory stimuli, that are otherwise obscured [79]. Taken together, these studies argue that accounting for movements will become increasingly important in decision-making studies.

**Movements and brain state**

Quantifying animal movements also gives experimenters a handle on tracking fluctuations in brain states (Table 1). Brain state changes account for large, widespread fluctuations in neural excitability, interneuronal correlations, oscillatory power of local-field potentials, stimulus-response amplitude and task performance [8,9,80–84]. Moreover, brain states can shape the transformation of sensory input to behavioral output. In fruit flies, similar sensory stimuli evoke different behaviors depending on brain state, where each state is associated with a unique set of behavioral output probabilities [75]. Despite their pronounced impact on neural activity and behavior the full range of distinct brain states is not fully characterized, especially in awake animals; this remains an area of active research [83].

Automatic recognition of behavioral motifs (see above) can be leveraged to continuously infer states like fear [85] or stress [86,87] or detect transitions from one state to another [75]. During decision-making, such behavioral quantification might aid in interpreting changes in choice behavior and neural activity on multiple timescales. For example, an animal might be stressed from exposure to a novel environment in the early stages of learning, but then habituate over time. Corresponding behavioral changes might be correlated with task performance and drive changes in neural activity that could be confounded with task learning. Continuous behavioral tracking would be a powerful way to address this concern.
A particularly valuable (involuntary) movement is the dilation of the pupil [8,9,82,84,88,89]. Fluctuations in pupil diameter are linked to release of acetylcholine and noradrenaline, with pupil dilations below 0.3 Hz being more closely related to cholinergic release and higher frequencies to noradrenergic release [90]. The interpretability of pupil diameter data might therefore be increased by using spectral analyses to isolate different neuromodulatory components that affect cortical processing. Phasic pupil-linked arousal can also reflect and interact with latent behavioral variables: for example, pupil dilation scales with decision uncertainty and predicts a reduction in across-trial serial choice biases in humans [91]. Combining behavioral models with continuous brain state measures therefore improves latent behavioral variable estimates and further increases their utility to reveal their underlying neural dynamics (Figure 2).

Locomotion also has profound effects on neural activity and sensory perception [10,83,92]. While sometimes used as an alternate measure of arousal in mice [10,92,93], locomotion is mainly associated with a high arousal state and does not accurately reflect the full spectrum of state modulations seen with pupil measures [9,83]. Pupil dilations are also seen during quiescence and can be observed seconds before and after bouts of locomotion [9,90]. Accordingly, locomotion is usually followed by severe depolarization or hyperpolarization across cortical areas for prolonged periods of time [93]. However, many sensory neurons are modulated by locomotion alone, independent of brain state. Auditory neurons are inhibited by locomotion, likely to suppress the perception of predictable, movement-related sounds [94] and while firing rates of V1 neurons are suppressed during arousal, they are elevated during locomotion [9]. The link between locomotion and changes in neural activity may also differ across species. In humans, locomotion does not affect neural responses or perception of visual stimuli, presumably because instructed walking in humans does not reflect the same arousal state as spontaneous running in mice [92].

Relating rich behavior to neural activity by studying ANNs
A new approach to relate neural dynamics to behavior is the use of ANNs as artificial model organisms. ANNs combine simple, nonlinear computational units connected together with adjustable parameters, in direct analogy to the neurons and synapses in biological brains [95,96]. Recurrent neural networks (RNNs) also contain recurrent feedback connections, again in analogy with brain anatomy. Usually, ANN parameters are adjusted through an iterative learning algorithm, called backpropagation, gradually improving performance over many behavioral trials. If given enough training examples, an ANN can perform extremely well at a task, often with dramatically improved performance over hand-designed solutions.

We view ANNs as artificial model organisms because after training a researcher can analyze the ANN to glean how the task was solved by the network. The ANN can be studied in terms of its behavior, network activations, and the trained parameter values (its connectome). An advantage of artificial over biological model organisms is that one can easily train thousands of networks while systematically varying the behavioral task, neural architecture and cost functions, thus enabling the study of large ensembles of potentially different solutions to a given behavioral task. One may also study learning in ANNs, for example by studying the evolution of the ANN’s parameters (the weights of connections in the network) through training or by equipping the ANN with more biophysically motivated training methods.

For reasons that remain to be fully theoretically understood, neural activations in ANNs share important features with those found in biological organisms [56,97], especially at the neural population level. This enables researchers to use ANNs to create and study two kinds of hypotheses concerning biological neural data. The first use of ANNs has been to generate hypotheses for whether neural activity might subserve an animal’s behavior [4,50–55]. If the internal population dynamics of a trained ANN can explain a large amount of variance in animal neural recordings, the cost function used to train the ANN is likely to be related to the animal’s task performance (Figure 2). For example, ANNs were used to discover that grid cell representations of space arise naturally as a solution to solving the problem of path integration [98,99]. In the second use of ANNs, one studies how the ANN dynamics implement its cost function. This is tantamount to reverse engineering the ANN, and if successful this approach can be used to yield novel hypotheses about how neural dynamics might support the animals’ behaviors (Figure 2). For example, brief oscillatory dynamics were discovered as a means of generating muscle activity from instructed reaches in a trained ANN [51].

A key advantage of ANNs over traditional latent variable models is the ability of ANNs to embrace the complexity and heterogeneity of rich experimental tasks and their resulting behaviors [97]. For example, traditional ideal observer models are often expressed as probabilities on how behavior is expected to relate to latent variables. These equations are then inverted using Bayes rule to recover the probabilities of the latent variable being studied. While simple and often elegant, ideal observer models (and cognitive models more generally) are unlikely to be able to explain the rich set of dynamics often seen in neural recordings, due to a desire to keep the model analytically tractable (though
hybrid approaches exist [21]). Additionally, cognitive models explicitly express a hypothesis for how a computation is performed, whereas ANNs are constructed to solve a loss function using neural dynamics and thus are hypothesis generators requiring reverse engineering. The richness of ANN modeling comes at the cost of then having to study a more complex model, but increased task complexity and more detailed behavioral measures can create new ways to better constrain ANN solutions. As technology matures, we expect ANNs to play an increasing role in revealing how complex neural dynamics give rise to rich behavior.

A related approach is to model more ancillary behavioral data as output for the ANN, for example, not just binary choices, but estimates of the animal’s brain state or body movements. For instance, instead of producing a binary decision, ANNs can produce complex motor outputs resembling behavioral motifs seen in animals [51]. It is likely that considering these kinds of behavioral details will bring the ANNs into further alignment with neural data. Finally, animals may be asked to perform an isolated behavior in the laboratory, but of course, the animal is performing many ongoing behaviors and the brain must support all of them. It is likely that studies of ANNs trained on multiple tasks could enrich the solutions for individual tasks and favor implementations that are also seen in living neural networks [100]. For example, Yang and colleagues studied how a single network implements a large [61**] or huge [101] number of cognitive and memory-based tasks and found functionally specific clusters for different cognitive processes, resembling cognitive specificity of neurons in prefrontal cortex.

There are some profound conceptual differences between animals and ANNs that impact behavioral modeling and will require additional research on the theory side. One large difference is that ANNs are trained only once during an optimization process and the connection weights are not subsequently modified, while animals continually update and refine their behavior. This discrepancy seems fine for understanding ‘instantaneous snapshots’ of animal behavior but is highly problematic for understanding how animals learn or how their neural representations evolve over time [100,102]. A related consideration is that biological brains implement both the computation underlying behavior as well as the system that enables learning of novel behaviors. ANNs, however, use externally available cost functions and optimization routines, typically written as auxiliary software, which are discarded after training. The incorporation of reinforcement learning to flexibly train ANNs might be a way to overcome this limitation and allow ANNs to uncover variable task contingencies on their own [100,103,104]. Extending this approach to large numbers of ANNs will likely enable the study of differing behavioral strategies as found in behaving animals.

Individual differences – from averages to individuality

Distinct behavioral strategies are part of a pervasive feature in many experimental and natural behaviors: individual differences (Table 1). Individuality refers to specific behavioral traits that differ across animals and impact their responsiveness to the environment. For example, even animals with a similar genetic background respond differently to pharmacological interventions [105] and stress [106], and display idiosyncratic behavioral strategies during decision-making [14**,42,63**,68].

Recognizing the role of individual behavioral strategy can significantly change the interpretation of neural recordings and perturbations during decision-making. Mice discriminating textures show activation of different cortical areas corresponding to distinct active or passive movement strategies [106]. Consequently, cortical inactivation only affects behavior underlying the corresponding strategy. Behavioral traits also vary over an animal’s life time: changing the social environment reshapes activity of midbrain dopamine neurons and animal strategy during foraging [68].

A recent success in relating individual choice behavior of rats in an auditory discrimination task to neural activity came from RNNs [14*]. Here, the internal dynamics of different random RNN instantiations matched neural dynamics from medial frontal cortex recordings across different animals, with choice selectivity emerging more strongly in high-performing rats and RNNs. Intriguingly, the stability of the RNN’s internal dynamics could also be measured without sensory stimulation, providing insights beyond the recorded neural activity [14*]. Future efforts will show if individual behaviors can be fit with custom ANN architectures [107], rather than using random network instantiations, to create models for individual animals’ neural and cognitive dynamics.

Conclusion

We have described how complex task design and in-depth behavioral quantification can be leveraged to gain insights into the interplay between behavioral output and the underlying neural activity. Rather than trying to constrain behavior and focusing on a few instructed movements in simple tasks, we argue that future studies will strongly benefit from embracing diversity as a feature of animal behavior to understand previously unknown features of neural activity. More detailed behavioral information should also extend into the realm of ANNs and might provide new ways to create stronger links between artificial and biological neural networks. A new experimental approach might incorporate both the task-based and observational approach to estimate latent variables and interpret neural data, while leveraging ANNs to make predictions about concomitant neural dynamics (Figure 2).
Conflict of interest statement
Nothing declared.

Acknowledgements
We thank Sashank Pisupati, Adam Calhoun and Ashley Juavinett for thoughtful conversations and advice on earlier versions of the manuscript. This work was supported by the Simons Collaboration on the Global Brain, Army Research Office (ARO) under contract W911NF-16-1-0368, the Pew Charitable Trust, the Eleanor Schwartz Trust, the Leopoldina National Academy of Science and the Swiss National Science Foundation.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

9. Brain state changes in mice were measured through self-initiated locomotion and pupil size. Pupil-linked state transitions were present in quiescent episodes and had effects on neural activity in visual cortex that were different from locomotion. Pupil dilations coincided with suppressed firing rates and enhanced visual responses while locomotion resulted in an overall increase in firing rates.


Recurrent ANNs were trained to perform a heat gradient navigation task, based on behavioral swimming patterns of zebrafish in response to temperature gradients. Clusters of ANN units showed response profiles similar to known cell types within the zebrafish hindbrain. Intriguingly, ANN activation clusters suggested an additional functional class of neurons, which was found upon closer inspection of calcium imaging data. Overall, the ‘fish-like’ ANN units were necessary for successful navigation behavior, whereas removing other ANN units did not affect the network’s performance.


Single RNNs were trained to perform 20 different cognitive tasks. After training, recurrent units formed clusters that were specifically tuned for different cognitive processes. The RNNs also showed mixed task selectivity, where some tasks could be solved by combining other task instructions. Training RNNs sequentially substantially increased such mixed task representation and resembled neural responses in monkey prefrontal cortex.


Animal movements during a texture-discrimination task reflected an active/passive strategy, leading to remarkably different cortical activity patterns in a subsequent delay period. Correspondingly, optogenetic inactivation of cortical areas had highly variable effects on task performance that were explained by the animal’s behavioral strategy.


This toolbox for automated animal pose tracking requires a hand-labeled training set (~100 frames) to achieve high prediction performance in flies or mice. Unsupervised dimensionality reduction of movement trajectories and subsequent clustering could identify ~20 distinct behavioral motifs of fruit fly behavior, for example, different grooming patterns.


A combination of hidden-Markov-model and generalized linear models was used to assess how different brain states shape sensorimotor transformation. In fruit fly experiments, the model accurately predicted behavioral state transitions, based on song history and various behaviors that were automatically extracted from video data. In each state, similar sensory inputs drove different behaviors, demonstrating that sensorimotor transformation is flexible and state- dependent.


