

Harnessing behavioral diversity to understand neural computations for cognition

Abstract

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.

Introduction

To understand the computations implemented by neural network dynamics, it is critical 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 [8–13].

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 (Fig. 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 (Fig. 1) [14].



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.

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 [22], 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?

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 [37–39]. Continuous responses like reaching movements, moving a wheel or navigating in virtual reality can provide additional insight into the evolving decision process [29,40–42]. For example, changes in head orientation are related to upcoming animal choices and trial-to-trial neural variability [43]. Further, reaching trajectories in a virtual maze task can reveal changes of mind in free-choice trials [29].

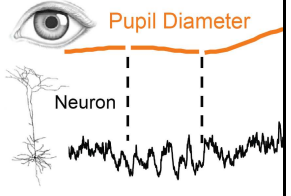
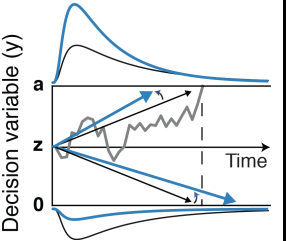
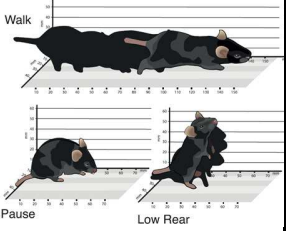

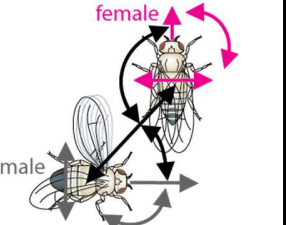
As animals naturally integrate information across many trials to combine sensory information with prior knowledge and beliefs [44], variations in task history can reveal the algorithms by which agents integrate information across trials [45], learn the structure of their environment [46], perform probabilistic inference [47], and adjust their decision policy after errors [48]. 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 [49] and single cells [50]. 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 [51].


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 [4,52–57]. 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 [58]. ANNs may also complement traditional behavioral models to capture the complexity of richer or more ethological tasks [59].

Box (separate) -----

How can rich behavioral tasks, combined with increasingly high-dimensional neural measurements, help us to relate complex neural dynamics to behavior? The theory of neural task complexity (NTC) states that the dimensionality of neural population dynamics has an upper bound defined by the number of task parameters and the smoothness of neural trajectories across those parameters [60]. In simple tasks, neural network dynamics are therefore constrained by the low number of task parameters and should contain far fewer dimensions compared to the number of recorded neurons. To overcome this issue, NTC can be used to compute the expected dimensionality of neural network dynamics when increasing task complexity. This framework thus promises to be a valuable tool to titrate the complexity required for future behavioral tasks to match our growing capacity for recording many neurons simultaneously [6].

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

Term	Definition	Examples	Biological vs artificial networks?	
Brain state	Internally generated neural dynamics that fluctuate spontaneously, often in ways that are related to bodily constraints. These are often measured through physiological markers.	Arousal, fear, stress, hunger, motivation, engagement, drowsiness	Biological:[61]	
Latent behavioral variable	The agent's estimate of a computationally relevant quantity. These latent variables are inferred via behavioral models.	Accumulated evidence, bias, value, confidence	Both (biological:[22] ANN:[21])	
Behavioral motif	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.	Grooming, eating, walking, running, jumping, rearing, climbing, reaching,	Both (biological:[62] ANN:[53])	
Task-based approach	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.	Random dot motion task, reaching task, maze navigation task, image classification task	Both (biological:[34] ANN:[63])	
Observational approach	Experiments where self-generated behavior of untrained animals is observed and analyzed.	Home cage exploration, mating behavior, head-fixed wheel locomotion, place field mapping	Biological:[64] (currently)	

Individual differences	Differences in an individual's behavioral repertoire that allow to distinguish different animal types.	Exploratory vs. fearful, dominant vs. submissive, social vs. asocial, active vs. inactive	Both (biological:[65] ANN:[14])	
------------------------	--	---	---------------------------------	---

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 [66]. 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 [64]. With the rise of neural recordings in freely-moving animals, such behavioral motifs are now routinely related to neural activity patterns [67,68].

There is an extensive toolbox for characterizing movements. This includes audio recording [69] and RFID tagging [70], 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 [71,72]. 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* [73,74], larval zebrafish [75], and *drosophila melanogaster* [76,77], along with freely moving mice [62]. Here, time-frequency analysis or auto-regressive models can discover and quantify the occurrence of stereotypic 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 [62,78], 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 [79].

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 over-constrained 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 [7,80], it is intriguing that highly-trained, head-fixed animals are far from being ‘task automatons’. 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 [8–13]. 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 don’t 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 [81]. 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 [8,9,82–86]. 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 [77]. 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 [85].

Automatic recognition of behavioral motifs (see above) can be leveraged to continuously infer states like fear [87] or stress [88,89] or detect transitions from one state to another [77]. 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 [8,9,84,86,90,91]. 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 [92]. 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 [93]. 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 [10,85,94]. While sometimes used as an alternate measure of arousal in mice [10,94,95], locomotion is mainly associated with a high arousal state and does not accurately reflect the full spectrum of state modulations seen with pupil measures [9,85]. Pupil dilations are also seen during quiescence and can be observed seconds before and after bouts of locomotion [9,92]. Accordingly, locomotion is usually followed by severe de- or hyperpolarization across cortical areas for prolonged periods of time [95]. 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 [96] 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 [94].

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 [97,98]. 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 back-propagation, 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 critical features with those found in biological organisms [58,99], 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 [4,52–57]. 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 [100,101]. 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 [53].

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 [99]. 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, e.g. 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, e.g. 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 [53]. 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 [102]. For example, Yang and colleagues studied how a single network implements a large [63] or huge [103] 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 [102,104]. 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 [102,105,106]. 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 [107] and stress [108], and display idiosyncratic behavioral strategies during decision-making [14,44,65,70].

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 [108]. 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 [70].

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 [109], 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).

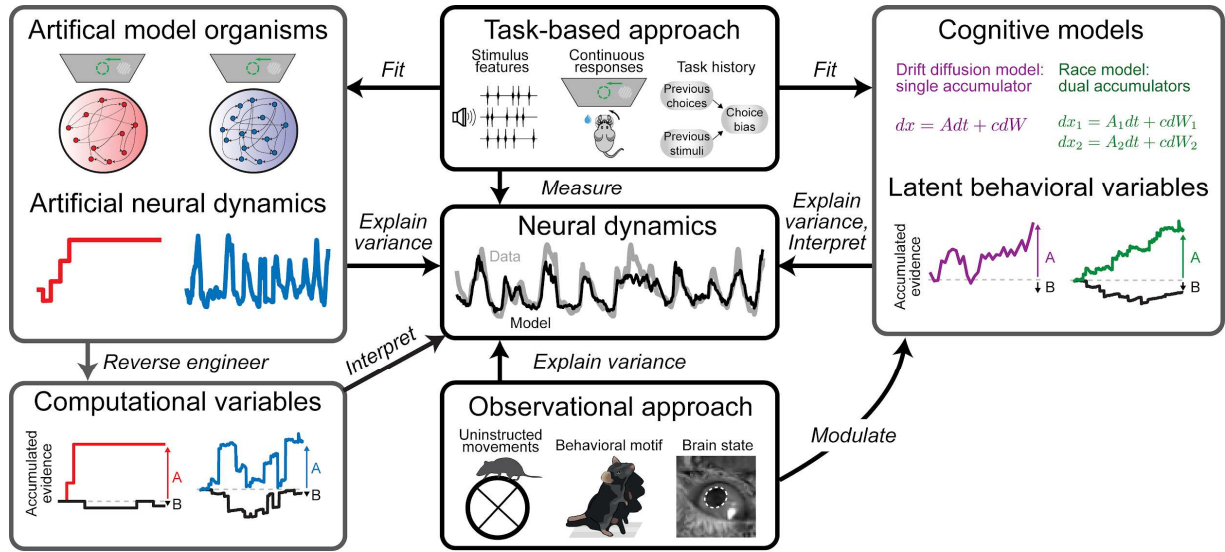


Figure 2. A candidate workflow schematizing how new experimental approaches (black boxes) and analyses (grey boxes) can be combined to gain insight into neural computations that underlie behavior. A **task-based approach** is used to generate rich behavioral and neural data about the animal's task performance. Workflow with **cognitive models**, (right): **Latent behavioral variables** are estimated from different task features, using explicit models of cognitive computations. An **observational approach** measures ongoing movements and brain state. Variability in **neural dynamics** is explained by combining movements, brain state and latent behavioral variables. Workflow with ANNs (left): multiple ANNs are trained on the same task, using different cost functions and hyperparameters, such as network connectivity or those controlling the complexity of the dynamics. ANNs function as **artificial model organisms** with slight variations to discover possible network implementations that support a cognitive computation. The scientist has access to the underlying artificial neural dynamics and ANN parameters, from which computational variables can be reverse-engineered to interpret an animal's neural data. These **computational variables** can be similar to those derived from cognitive models, but apply to a broader class of tasks where explicit cognitive models may not be available or tractable.

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, ONR MURI, the Pew Charitable Trust, the Eleanor Schwartz Trust, the Leopoldina National Academy of Science and the Swiss National Science Foundation.

Conflict of interest statement: Nothing declared.

Highlighted references

Vinck et al., 2015 *:

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.

Musall et al., 2019 *:

A linear regression model combined video-based movement representations and task variables to predict changes in cortex-wide neural activity during decision-making. Uninstructed movements accounted for a large fraction of neural variance, outperforming instructed movements or task variables. Model predictions accounted for a large amount of trial-to-trial variability and could be used to identify the impact of movements on trial-averaged data.

Gilad et al., 2018 **:

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.

Kurikawa et al., 2018 *:

Rats in an auditory detection task exhibited individual variability in their responses to unfamiliar stimuli, reflected in the stability of neural dynamics in medial frontal cortex (MFC). Multiple RNN models of MFC with randomly initiated recurrent connections recapitulated these individual behavioral and neural patterns.

Pereira et al. 2019 *:

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, e.g. different grooming patterns.

Yang et al. 2019 **:

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.

Calhoun, Pillow and Murthy 2019 **

A combination of hidden-markov-model and generalized linear models was used to assess how different brain states shape sensorimotor transformation. In fruit fly courtship, 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.

Haesemeyer et al. 2019 **

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.

References

1. Uchida N, Mainen ZF: **Speed and accuracy of olfactory discrimination in the rat.** *Nat Neurosci* 2003, **6**:1224–1229.
2. Freedman DJ, Assad JA: **Experience-dependent representation of visual categories in parietal cortex.** *Nature* 2006, **443**:85–88.
3. Raposo D, Sheppard JP, Schrater PR, Churchland AK: **Multisensory decision-making in rats and humans.** *J Neurosci Off J Soc Neurosci* 2012, **32**:3726–3735.
4. Mante V, Sussillo D, Shenoy KV, Newsome WT: **Context-dependent computation by recurrent dynamics in prefrontal cortex.** *Nature* 2013, **503**:78–84.
5. Gomez-Marín A, Paton JJ, Kampff AR, Costa RM, Mainen ZF: **Big behavioral data: psychology, ethology and the foundations of neuroscience.** *Nat Neurosci* 2014, **17**:1455–1462.
6. Gao P, Ganguli S: **On simplicity and complexity in the brave new world of large-scale neuroscience.** *Curr Opin Neurobiol* 2015, **32**:148–155.
7. Krakauer JW, Ghazanfar AA, Gomez-Marín A, MacIver MA, Poeppel D: **Neuroscience Needs Behavior: Correcting a Reductionist Bias.** *Neuron* 2017, **93**:480–490.
8. McGinley MJ, David SV, McCormick DA: **Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection.** *Neuron* 2015, **87**:179–192.
9. Vinck M, Batista-Brito R, Knoblich U, Cardin JA: **Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding.** *Neuron* 2015, **86**:740–754.
10. Niell CM, Stryker MP: **Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex.** *Neuron* 2010, **65**:472–479.
11. Crochet S, Petersen CCH: **Correlating whisker behavior with membrane potential in barrel cortex of awake mice.** *Nat Neurosci* 2006, **9**:608–610.
12. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD: **Spontaneous behaviors drive multidimensional, brainwide activity.** *Science* 2019, **364**:eaav7893.
13. Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK: **Single-trial neural dynamics are dominated by richly varied movements.** *bioRxiv* 2019, doi:10.1101/308288.
14. 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.
15. Kriegeskorte N, Douglas PK: **Cognitive computational neuroscience.** *Nat Neurosci* 2018, **21**:1148–1160.
16. Ratcliff R, McKoon G: **The diffusion decision model: theory and data for two-choice decision tasks.** *Neural Comput* 2008, **20**:873–922.
17. Shadlen MN, Kiani R: **Decision making as a window on cognition.** *Neuron* 2013, **80**:791–806.
18. Sutton R, Barto A: **Reinforcement Learning.** *MIT Press* 1998.
19. Dayan P, Daw ND: **Decision theory, reinforcement learning, and the brain.** *Cogn Affect Behav Neurosci* 2008, **8**:429–453.
20. Ma WJ: **Organizing probabilistic models of perception.** *Trends Cogn Sci* 2012, **16**:511–518.
21. Echeveste R, Aitchison L, Hennequin G, Lengyel M: **Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference.** *bioRxiv* 2019, doi:10.1101/696088.
22. Gold JI, Shadlen MN: **The Neural Basis of Decision Making.** *Annu Rev Neurosci* 2007, **30**:535–574.
23. Fiorillo CD, Tobler PN, Schultz W: **Discrete Coding of Reward Probability and**

- Uncertainty by Dopamine Neurons.** *Science* 2003, **299**:1898–1902.
24. Churchland AK, Kiani R, Shadlen MN: **Decision-making with multiple alternatives.** *Nat Neurosci* 2008, **11**:693–702.
 25. Schall JD: **On building a bridge between brain and behavior.** *Annu Rev Psychol* 2004, **55**:23–50.
 26. Evans DA, Stempel AV, Vale R, Ruehle S, Lefler Y, Branco T: **A synaptic threshold mechanism for computing escape decisions.** *Nature* 2018, **558**:590–594.
 27. McDougle SD, Butcher PA, Parvin DE, Mushtaq F, Niv Y, Ivry RB, Taylor JA: **Neural Signatures of Prediction Errors in a Decision-Making Task Are Modulated by Action Execution Failures.** *Curr Biol CB* 2019, **29**:1606-1613.e5.
 28. Donner TH, Siegel M, Fries P, Engel AK: **Buildup of choice-predictive activity in human motor cortex during perceptual decision making.** *Curr Biol CB* 2009, **19**:1581–1585.
 29. Kaufman MT, Churchland MM, Ryu SI, Shenoy KV: **Vacillation, indecision and hesitation in moment-by-moment decoding of monkey motor cortex.** *eLife* 2015, **4**:e04677.
 30. Hanks TD, Kopec CD, Brunton BW, Duan CA, Erlich JC, Brody CD: **Distinct relationships of parietal and prefrontal cortices to evidence accumulation.** *Nature* 2015, **520**:220–223.
 31. Yates JL, Park IM, Katz LN, Pillow JW, Huk AC: **Functional dissection of signal and noise in MT and LIP during decision-making.** *Nat Neurosci* 2017, **20**:1285–1292.
 32. Brunton BW, Botvinick MM, Brody CD: **Rats and Humans Can Optimally Accumulate Evidence for Decision-Making.** *Science* 2013, **340**:95–98.
 33. Kiani R, Hanks TD, Shadlen MN: **Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment.** *J Neurosci Off J Soc Neurosci* 2008, **28**:3017–3029.
 34. Sheppard JP, Raposo D, Churchland AK: **Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans.** *J Vis* 2013, **13**.
 35. Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE: **Neural correlates of reliability-based cue weighting during multisensory integration.** *Nat Neurosci* 2011, **15**:146–154.
 36. Roitman JD, Shadlen MN: **Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task.** *J Neurosci* 2002, **22**:9475–9489.
 37. Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepecs A: **Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence.** *Neuron* 2014, **84**:190–201.
 38. Lak A, Nomoto K, Keramati M, Sakagami M, Kepecs A: **Midbrain Dopamine Neurons Signal Belief in Choice Accuracy during a Perceptual Decision.** *Curr Biol CB* 2017, **27**:821–832.
 39. Fleming SM, Putten EJ van der, Daw ND: **Neural mediators of changes of mind about perceptual decisions.** *Nat Neurosci* 2018, **21**:617.
 40. Burgess CP, Steinmetz N, Lak A, Zatzka-Haas P, Ranson A, Wells M, Schroeder S, Jacobs EAK, Reddy CB, Soares S, et al.: **High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice.** *bioRxiv* 2016, doi:10.1101/051912.
 41. Morcos AS, Harvey CD: **History-dependent variability in population dynamics during evidence accumulation in cortex.** *Nat Neurosci* 2016, advance online publication.
 42. Huk A, Bonnen K, He BJ: **Beyond Trial-Based Paradigms: Continuous Behavior, Ongoing Neural Activity, and Natural Stimuli.** *J Neurosci* 2018, **38**:7551–7558.
 43. Erlich JC, Bialek M, Brody CD: **A Cortical Substrate for Memory-Guided Orienting in**

- the Rat.** *Neuron* 2011, **72**:330–343.
44. Urai AE, de Gee JW, Tsetsos K, Donner TH: **Choice history biases subsequent evidence accumulation.** *eLife* 2019, **8**:e46331.
 45. Glaze CM, Kable JW, Gold JL: **Normative evidence accumulation in unpredictable environments.** *eLife* 2015, **4**:e08825.
 46. Miller KJ, Botvinick MM, Brody CD: **From predictive models to cognitive models: An analysis of rat behavior in the two-armed bandit task.** *bioRxiv* 2019, doi:10.1101/461129.
 47. Drugowitsch J, Wyart V, Devauchelle A-D, Koechlin E: **Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality.** *Neuron* 2016, **92**:1398–1411.
 48. Purcell BA, Kiani R: **Neural Mechanisms of Post-error Adjustments of Decision Policy in Parietal Cortex.** *Neuron* 2016, **89**:658–671.
 49. O'Doherty JP, Hampton A, Kim H: **Model-based fMRI and its application to reward learning and decision making.** *Ann N Y Acad Sci* 2007, **1104**:35–53.
 50. Fernandes HL, Stevenson IH, Phillips AN, Segraves MA, Kording KP: **Saliency and saccade encoding in the frontal eye field during natural scene search.** *Cereb Cortex N Y N 1991* 2014, **24**:3232–3245.
 51. Roy NA, Bak JH, Akrami A, Brody C, Pillow JW: **Efficient inference for time-varying behavior during learning.** In *Advances in Neural Information Processing Systems 31*. Edited by Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R. Curran Associates, Inc.; 2018:5695–5705.
 52. Carnevale F, de Lafuente V, Romo R, Barak O, Parga N: **Dynamic Control of Response Criterion in Premotor Cortex during Perceptual Detection under Temporal Uncertainty.** *Neuron* 2015, **86**:1067–1077.
 53. Sussillo D, Churchland MM, Kaufman MT, Shenoy KV: **A neural network that finds a naturalistic solution for the production of muscle activity.** *Nat Neurosci* 2015, **18**:1025–1033.
 54. Rajan K, Harvey CD, Tank DW: **Recurrent Network Models of Sequence Generation and Memory.** *Neuron* 2016, **90**:128–142.
 55. Remington ED, Narain D, Hosseini EA, Jazayeri M: **Flexible Sensorimotor Computations through Rapid Reconfiguration of Cortical Dynamics.** *Neuron* 2018, **98**:1005–1019.e5.
 56. Sohn H, Narain D, Meirhaeghe N, Jazayeri M: **Bayesian computation through cortical latent dynamics.** *bioRxiv* 2018, doi:10.1101/465419.
 57. Wang J, Narain D, Hosseini EA, Jazayeri M: **Flexible timing by temporal scaling of cortical responses.** *Nat Neurosci* 2018, **21**:102–110.
 58. Haesemeyer M, Schier AF, Engert F: **Convergent Temperature Representations in Artificial and Biological Neural Networks.** *Neuron* 2019, **0**.
 59. Yamins DLK, DiCarlo JJ: **Using goal-driven deep learning models to understand sensory cortex.** *Nat Neurosci* 2016, **19**:356–365.
 60. Gao P, Trautmann E, Yu BM, Santhanam G, Ryu S, Shenoy K, Ganguli S: **A theory of multineuronal dimensionality, dynamics and measurement.** *bioRxiv* 2017, doi:10.1101/214262.
 61. McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, Cadwell CR, Tolias AS, Cardin JA, McCormick DA: **Waking State: Rapid Variations Modulate Neural and Behavioral Responses.** *Neuron* 2015, **87**:1143–1161.
 62. Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abaira VE, Adams RP, Datta SR: **Mapping Sub-Second Structure in Mouse Behavior.** *Neuron* 2015, **88**:1121–1135.
 63. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J: **Task representations in**

- neural networks trained to perform many cognitive tasks.** *Nat Neurosci* 2019, doi:10.1038/s41593-018-0310-2.
64. Anderson DJ, Perona P: **Toward a Science of Computational Ethology.** *Neuron* 2014, **84**:18–31.
 65. Gilad A, Gallero-Salas Y, Groos D, Helmchen F: **Behavioral Strategy Determines Frontal or Posterior Location of Short-Term Memory in Neocortex.** *Neuron* 2018, **99**:814–828.e7.
 66. Tinbergen N: *The Study of Instinct: with a new Preface.* Oxford University Press; 1989.
 67. Doupe AJ, Konishi M: **Song-selective auditory circuits in the vocal control system of the zebra finch.** *Proc Natl Acad Sci U S A* 1991, **88**:11339–11343.
 68. Stensmyr MC, Dweck HKM, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, Grabe V, Steck K, Lavista-Llanos S, et al.: **A conserved dedicated olfactory circuit for detecting harmful microbes in Drosophila.** *Cell* 2012, **151**:1345–1357.
 69. Coffey KR, Marx RG, Neumaier JF: **DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations.** *Neuropsychopharmacology* 2019, doi:10.1038/s41386-018-0303-6.
 70. Torquet N, Marti F, Campart C, Tolu S, Nguyen C, Oberto V, Benallaoua M, Naudé J, Didienne S, Debray N, et al.: **Social interactions impact on the dopaminergic system and drive individuality.** *Nat Commun* 2018, **9**:3081.
 71. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M: **DeepLabCut: markerless pose estimation of user-defined body parts with deep learning.** *Nat Neurosci* 2018, **21**:1281.
 72. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS-H, Murthy M, Shaevitz JW: **Fast animal pose estimation using deep neural networks.** *Nat Methods* 2019, **16**:117.
 73. Stephens GJ, Johnson-Kerner B, Bialek W, Ryu WS: **Dimensionality and Dynamics in the Behavior of C. elegans.** *PLOS Comput Biol* 2008, **4**:e1000028.
 74. Kato S, Kaplan HS, Schrödel T, Skora S, Lindsay TH, Yemini E, Lockery S, Zimmer M: **Global brain dynamics embed the motor command sequence of Caenorhabditis elegans.** *Cell* 2015, **163**:656–669.
 75. Girdhar K, Gruebele M, Chemla YR: **The Behavioral Space of Zebrafish Locomotion and Its Neural Network Analog.** *PloS One* 2015, **10**:e0128668.
 76. Berman GJ, Choi DM, Bialek W, Shaevitz JW: **Mapping the stereotyped behaviour of freely moving fruit flies.** *J R Soc Interface* 2014, **11**.
 77. Calhoun AJ, Pillow JW, Murthy M: **Unsupervised identification of the internal states that shape natural behavior.** *bioRxiv* 2019, doi:10.1101/691196.
 78. Markowitz JE, Gillis WF, Beron CC, Neufeld SQ, Robertson K, Bhagat ND, Peterson RE, Peterson E, Hyun M, Linderman SW, et al.: **The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection.** *Cell* 2018, **174**:44–58.e17.
 79. Clemens J, Coen P, Roemschied FA, Pereira TD, Mazumder D, Aldarondo DE, Pacheco DA, Murthy M: **Discovery of a New Song Mode in Drosophila Reveals Hidden Structure in the Sensory and Neural Drivers of Behavior.** *Curr Biol* 2018, **28**:2400–2412.e6.
 80. Juavinett AL, Erlich JC, Churchland AK: **Decision-making behaviors: weighing ethology, complexity, and sensorimotor compatibility.** *Curr Opin Neurobiol* 2018, **49**:42–50.
 81. Andrew Hires S, Gutnisky DA, Yu J, O'Connor DH, Svoboda K: **Low-noise encoding of active touch by layer 4 in the somatosensory cortex.** *eLife* 2015, **4**:e06619.
 82. Buzsáki G, Draguhn A: **Neuronal oscillations in cortical networks.** *Science* 2004, **304**:1926–1929.
 83. Fox MD, Snyder AZ, Vincent JL, Raichle ME: **Intrinsic Fluctuations within Cortical Systems Account for Intertrial Variability in Human Behavior.** *Neuron* 2007, **56**:171–

- 184.
84. Cohen MR, Maunsell JHR: **Attention improves performance primarily by reducing interneuronal correlations.** *Nat Neurosci* 2009, **12**:1594–1600.
85. Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, Huang ZJ, Stryker MP: **A Cortical Circuit for Gain Control by Behavioral State.** *Cell* 2014, **156**:1139–1152.
86. Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS: **Pupil fluctuations track fast switching of cortical states during quiet wakefulness.** *Neuron* 2014, **84**:355–362.
87. Pham J, Cabrera SM, Sanchis-Segura C, Wood MA: **Automated scoring of fear related behavior using EthoVision software.** *J Neurosci Methods* 2009, **178**:323–326.
88. Füzesi T, Daviu N, Wamsteeker Cusulin JI, Bonin RP, Bains JS: **Hypothalamic CRH neurons orchestrate complex behaviours after stress.** *Nat Commun* 2016, **7**:11937.
89. Egan RJ, Bergner CL, Hart PC, Cachat JM, Canavello PR, Elegante MF, Elkhayat SI, Bartels BK, Tien AK, Tien DH, et al.: **Understanding behavioral and physiological phenotypes of stress and anxiety in zebrafish.** *Behav Brain Res* 2009, **205**:38–44.
90. Hess EH, Polt JM: **Pupil size as related to interest value of visual stimuli.** *Science* 1960, **132**:349–350.
91. Hess EH, Polt JM: **Pupil Size in Relation to Mental Activity during Simple Problem-Solving.** *Science* 1964, **143**:1190–1192.
92. Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, Tolias AS: **Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex.** *Nat Commun* 2016, **7**.
93. Urai AE, Braun A, Donner TH: **Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias.** *Nat Commun* 2017, **8**:14637.
94. Polack P-O, Friedman J, Golshani P: **Cellular mechanisms of brain state-dependent gain modulation in visual cortex.** *Nat Neurosci* 2013, **16**:1331–1339.
95. Shimaoka D, Harris KD, Carandini M: **Effects of Arousal on Mouse Sensory Cortex Depend on Modality.** *Cell Rep* 2018, **22**:3160–3167.
96. Schneider DM, Sundararajan J, Mooney R: **A cortical filter that learns to suppress the acoustic consequences of movement.** *Nature* 2018, **561**:391.
97. Zipser D, Andersen RA: **A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons.** *Nature* 1988, **331**:679–684.
98. Fetz EE: **Are movement parameters recognizably coded in the activity of single neurons?** *Behav Brain Sci* 1992, **15**:679–690.
99. Maheswaranathan N, Williams AH, Golub MD, Ganguli S, Sussillo D: **Universality and individuality in neural dynamics across large populations of recurrent networks.** *ArXiv190708549 Cs Q-Bio* 2019,
100. Banino A, Barry C, Uria B, Blundell C, Lillicrap T, Mirowski P, Pritzel A, Chadwick MJ, Degris T, Modayil J, et al.: **Vector-based navigation using grid-like representations in artificial agents.** *Nature* 2018, **557**:429–433.
101. Cueva CJ, Wei X-X: **Emergence of grid-like representations by training recurrent neural networks to perform spatial localization.** *ArXiv180307770 Cs Q-Bio Stat* 2018,
102. Flesch T, Balaguer J, Dekker R, Nili H, Summerfield C: **Comparing continual task learning in minds and machines.** *Proc Natl Acad Sci* 2018, **115**:E10313–E10322.
103. Yang GR, Ganichev I, Wang X-J, Shlens J, Sussillo D: **A Dataset and Architecture for Visual Reasoning with a Working Memory.** *ArXiv180306092 Cs* 2018,
104. Driscoll LN, Pettit NL, Minderer M, Chettih SN, Harvey CD: **Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex.** *Cell* 2017, **170**:986–999.e16.
105. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D,

- Botvinick M: **Prefrontal cortex as a meta-reinforcement learning system**. *bioRxiv* 2018, doi:10.1101/295964.
106. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, et al.: **Human-level control through deep reinforcement learning**. *Nature* 2015, **518**:529–533.
107. Morgan D, Grant KA, Gage HD, Mach RH, Kaplan JR, Prioleau O, Nader SH, Buchheimer N, Ehrenkaufer RL, Nader MA: **Social dominance in monkeys: dopamine D2 receptors and cocaine self-administration**. *Nat Neurosci* 2002, **5**:169–174.
108. Russo SJ, Murrough JW, Han M, Charney DS, Nestler EJ: **Neurobiology of Resilience**. *Nat Neurosci* 2012, **15**:1475–1484.
109. Hassabis D, Kumaran D, Summerfield C, Botvinick M: **Neuroscience-Inspired Artificial Intelligence**. *Neuron* 2017, **95**:245–258.