



Toward the neural implementation of structure learning

D Gowanlock R Tervo¹, Joshua B Tenenbaum² and Samuel J Gershman³

Despite significant advances in neuroscience, the neural bases of intelligence remain poorly understood. Arguably the most elusive aspect of intelligence is the ability to make robust inferences that go far beyond one's experience. Animals categorize objects, learn to vocalize and may even estimate causal relationships — all in the face of data that is often ambiguous and sparse. Such inductive leaps are thought to result from the brain's ability to infer latent structure that governs the environment. However, we know little about the neural computations that underlie this ability. Recent advances in developing computational frameworks that can support efficient structure learning and inductive inference may provide insight into the underlying component processes and help pave the path for uncovering their neural implementation.

Addresses

¹ Janelia Research Campus, 19700 Helix Dr., Ashburn, VA 20147, USA

² Department of Brain and Cognitive Sciences, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

³ Department of Psychology and Center for Brain Science, Harvard University, Northwest Building, 52 Oxford Street, Cambridge, MA 02138, USA

Corresponding author: Tervo, D Gowanlock R (gowan@janelia.hhmi.org)

Current Opinion in Neurobiology 2016, **37**:99–105

This review comes from a themed issue on **Neurobiology of behavior**

Edited by **Alla Karpova** and **Roosbeh Kiani**

<http://dx.doi.org/10.1016/j.conb.2016.01.014>

0959-4388/© 2016 Elsevier Ltd. All rights reserved.

Introduction

Animals perceive complex objects, learn abstract concepts and acquire sophisticated motor skills, often from limited experience. Efficiently making these inferences is paramount for survival, such as when determining whether a looming shadow indicates a predator, learning when and where to re-hide a food cache, or deciding to abandon a rich foraging niche in anticipation of a natural calamity. Inferences from sparse data depend upon background knowledge that restricts the potentially unlimited ways of parsing and interpreting the world. The brain likely makes these inferences by efficiently exploiting regularities in the environment to learn and use latent structured

relations. In essence, these structures are possible generative models that capture, at an abstract level, the relationships and causal processes underlying observations. Learned structure constraints can then be applied to solve related but novel tasks, such as parsing ambiguous sensory input and generating novel actions. How hidden structure is learned and used to support inductive leaps that go beyond the available data is an important question in contemporary neuroscience.

For more than a century, the challenge of determining how structure learning and inductive inference can be efficiently performed was tackled by statisticians [1], linguists [2], computer scientists [3] and cognitive scientists [4–7]. However, insights about the neural implementation of structure learning have been rare despite some excellent attempts [8,9]. Presumably this is because it is challenging to design experiments with the necessary task complexity, to ascertain that animals acquire and use specific structures and then to probe the underlying neural computations of structure learning and use. Nevertheless, if the need to learn structures that can support inductive inference was a selective pressure in the evolution of neural circuit function, then probing neural dynamics in this computational regime may be important for our understanding of brain function. Here we provide an overview of a computational approach to structure learning and inductive inference from contemporary cognitive science, and discuss what this framework offers to studies of the neural implementation of structure learning.

Structure learning in animal cognition

To gain an intuition for the advantage of knowing the appropriate generative structure of the environment, consider an entertaining anecdote about Richard Feynman. While bored at Los Alamos when he worked on the Manhattan Project, Feynman passed his time by picking the locks on filing cabinets [10]. By tinkering with the sophisticated three-disk Mosler combination locks that in principle could support one million combinations, Feynman uncovered certain regularities in the locks' design that greatly constrained the space of working combinations. Using this knowledge, he entertained himself by breaking into the cabinets of his colleagues whenever he needed a particular document. A similar insight into hidden regularities was the basis of Alan Turing's code-breaking exploits during World War II.

Feynman's use of a few observations to discover constraints that solved an otherwise irresolvable problem is a

striking example of the power of the human mind to learn about the workings of the world that go beyond the available information. Although it is accepted that the ability to abstract structured relationships in the environment is at the core of human intelligence, whether animals perform similar cognitive feats or whether they are instead ‘competent without comprehending’ remains contentious [11,12]. Nevertheless, an influential view posits that these remarkable, domain-general human cognitive abilities arose as an elaboration of more primitive, perhaps domain-specific cognitive systems [7,13,14]. Consistent with this claim, analyses of complex domain-specific behaviors in expert animals — e.g. rapid acquisition of social hierarchies in fish [15]; numerosity, tool use and intuitive physics in non-human primates [16–19] and corvids [20–22]; causal reasoning in rodents [23]; spatial navigation using schemas of interconnected locations [24,25^{*}] or even full cognitive maps [14] in rodents; and recognition and use of natural or artificial syntax in songbirds [26,27] — have provided evidence of animals’ reliance on rich ‘internal’ models of environmental structure that may resemble hierarchical symbolic abstractions central to human cognition.

Because inferring a generative structure amounts to abstracting general principles from examples, a key prediction of structure learning is that learning similar tasks is facilitated — a phenomenon known as ‘transfer learning’ or ‘learning to learn’ [28]. For example, a neophyte violinist who is a proficient pianist will become a skilled string instrumentalist quicker than someone new to music. Although the required movements of playing a keyboard differ from bowing a violin, experienced musicians know the principles of meter, rhythm, melody and harmony that they transfer to playing novel instruments.

Transfer learning is a widespread test for determining whether an animal has acquired structured abstractions. In a recent study, that also attempted to find a representation of the inferred task structure in neural activity, McKenzie *et al.* trained rats to learn a rule [25^{*}]. The rats were trained that for any pair of presented objects (flower pots scented with different odors), only one is rewarded and that the identity of the rewarded object in a pair remains the same independent of their placement in a given arena but changes between arenas. McKenzie *et al.* found that learning which pot to approach with the first pair of odors, X and Y, took twice as long compared to the subsequent odor pairs of A and B, and C and D, indicating transfer learning. Although these results support the notion that model organisms induce the structure of behavioral tasks, caution is warranted. In principle, transfer learning can be achieved without induction of an abstract rule: a simpler solution, such as the learning of a good way to group perceptions, can facilitate performance in a similar but novel environment. Thus, the development of more sophisticated behavioral frameworks and formal

approaches to verifying the specific form of structured knowledge an animal uses to solve a task are important challenges for the field (see below).

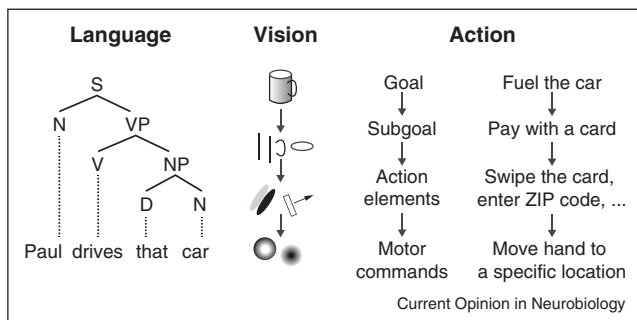
Modeling structure learning to generate insight into component processes

Any effort to understand the neural implementation of cognition requires that the underlying cognitive processes be identified and exposed in specifically-tailored behavioral tasks. These component cognitive processes can sometimes be intuited, such as evidence integration in perceptual judgments [29]. However, the algorithmic steps necessary for the acquisition of and reasoning with hierarchically structured abstractions of the environment are not immediately apparent. A rigorous approach has thus been developed within the broader framework of probabilistic models of cognition, which views human behavior in complex environments as solving a statistical inference problem. This approach starts with a *rational analysis* [30] of cognition — a normative characterization of how probabilistic reasoning *should* be performed given the available data and prior inductive constraints. The architecture of these normative characterizations specifies the requisite cognitive processes and their interactions, providing an explicit computational description of how cognition might be performed.

Within the probabilistic reasoning framework, nonparametric hierarchical Bayesian models (NPHBMs) incorporate the view that animals rely on structure learning, but accommodate uncertainty about the relationships and causal processes underlying observations by entertaining multiple (or even an infinite number of) candidate world structures. Most structured relationships in the environment can be represented as graphs, with entities or events modeled as nodes and relationships between them as edges, and the ability to organize knowledge accordingly is thought to be central to most of cognition (Figure 1). Notably, the proper form of the graph that best describes a particular latent structured relationship can be efficiently discovered by NPHBMs; the graph can also be grown, in an informal sense nonparametrically, to accommodate new observations.

For graphs, the NPHBM approach translates the subject’s uncertainty about the specific structure that best describes the observations into a space of discrete hypotheses at every level of abstraction. Interestingly, higher order abstractions can often be learnt before lower-level details, a phenomenon known as the *blessing of abstraction* [31,32]. For example, when children are presented with a few novel concepts and labels, they acquire a bias not only that individual categories like chairs tend to be organized by shape, but that categories of solid objects in general are as well [33]. One reason why higher-levels might be easier to learn is because the higher-level hypothesis space tends to be smaller and pools evidence

Figure 1



Nonparametric hierarchical Bayesian models are a popular method of growing and inferring generative structures of the world. Three example domains in which hierarchical structure supports generative or inferential processes (after [62]). In language sentences are parsed to semantic meaning, in vision whole objects can be decomposed into a hierarchy of progressively more primitive parts and in action selection the specific sequence of movements that lead to a distant goal can be planned. These hierarchies can be generative, such as the planning of a movement sequence, or inferential, such as the inference of the probable goal of an observed movement sequence.

from lower levels. Once higher abstraction is acquired, the lower-level hypothesis space becomes constrained, permitting rapid learning about related, but novel situations.

Probabilistic inference over structured forms thus permits knowledge to be simultaneously acquired at multiple levels and distributed according to the known relationships between entities or events. This is further highlighted by the ability of NPHBMs to provide a normative account for how new, sparsely observed properties should be generalized. For example, learning that a heritable disease afflicts a particular family member enables — via belief propagation over the ancestry tree — inference about other relatives at risk for the same illness. Overall, the power of NPHBMs compared with other contemporary computational accounts of cognition thus stems from their ability to discover the proper form of structured relationships, to infer the specific structure for a given set of observations, and to make predictions about how to generalize new, sparsely observed properties — all through rational statistical inference.

Probabilistic models of structured relationships are an attractive solution to the problem of structure learning because they allow multiple hypotheses to be considered in parallel. However, as the complexity of the task and the number of observations increases, the space of possible structures quickly explodes, making the search for the right one intractable. One solution to this issue is to use prior knowledge to constrain the hypothesis space [34]. For example, the existence of a prior expectation for which temporal scale is relevant for causality can greatly

reduce the number of hypothetical cause-effect relationships entertained by the model, and recent experimental evidence provides support for the existence of such priors in humans [35]. More generally, making the search for the right structure tractable tends to require a combination of smart priors on whatever relevant aspects of the structure the animal knows about, and smart search algorithms, which do not exhaustively search all possible structures but follow some efficient approximate search ordering [36].

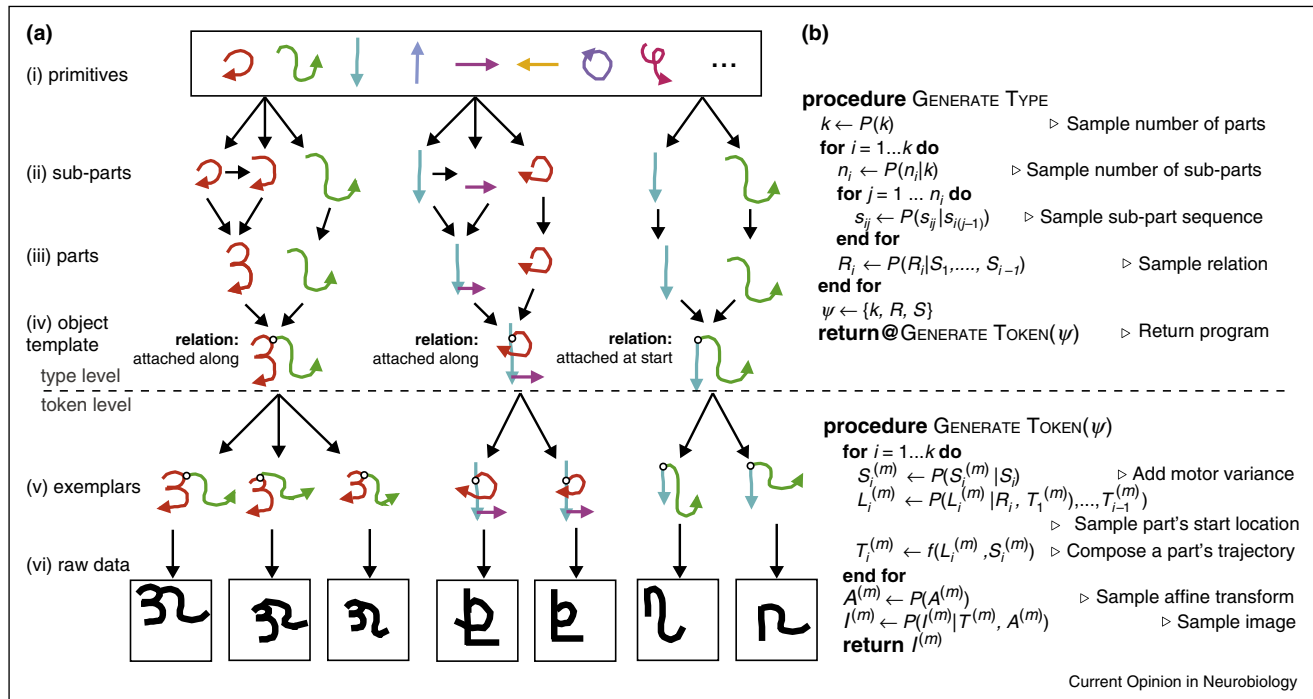
NPHBMs have gained particular recognition as normative accounts over the past decade due to their ability to capture some of the most impressive feats of human [5] and animal [37] cognition. In a recent example, Lake *et al.* (2015) compared the ability of NPHBMs and state-of-the-art artificial neural network approaches to humans' ability to grasp novel concepts — specifically, unfamiliar characters [38**] (Figure 2). Using a dataset of a handful of examples each for 1623 different characters — exactly the data-limited case that mandates the sharing of statistical information across exemplars — Lake *et al.* demonstrated that NPHBMs can both recognize and generate characters on par with humans. Central to their model's success was its ability to discover latent causal relations in the stroke-by-stroke process of generating written characters and compose these causal processes to recognize and generate new characters. This and other examples highlight the power of combining structure learning and statistical inference to enable such feats of cognition as transfer learning, one-shot learning and compositionality.

Toward the neural implementation of structure learning

NPHBMs have achieved human-level performance on a range of cognitive tasks including acquisition of novel concepts [38**], causal learning [39], parsing motion in the environment [40] and others (for a review see [5]). Although the extent to which this normative approach provides an adequate framework for cognition continues to be debated [41], the utility of such an abstract framework for systems neuroscience may come from the insights it offers into the component processes that might underlie the learning and use of latent structure.

Is it plausible that neural circuits implement approximations of the computational components of NPHBMs: structured hierarchies that can be grown nonparametrically, probabilistic inference over structured representations, and prior constraints on the space of structures an animal might consider? The hierarchical design of neural circuits, and the ability of animals to learn novel hierarchical abstractions of their environment has long been appreciated by neuroscientists [42,43] and is becoming an area of growing interest [44]. Of particular note, action planning in complex environments is increasingly thought to depend on hierarchical processing of subgoals

Figure 2



A generative model of handwritten characters that demonstrates the power of nonparametric hierarchical Bayesian models (reproduced from [38**]).

(a) New character types are generated by choosing primitive actions (color coded) from a library (i), combining these subparts (ii) to make parts (iii), and combining parts with relations to define simple programs (iv). New character tokens are generated by running these programs (v), which are then rendered as raw data (vi). **(b)** Pseudocode for generating new types ψ and new token images $I^{(m)}$ for $m = 1, \dots, M$. The function $f(\cdot, \cdot)$ transforms a subpart sequence and start location into a trajectory.

that can be separated significantly in time (for an excellent review see [45]) — although identifying the exact structure of subgoals in a particular task remains a challenge even with great control over experimental design [46]. Importantly, the existing experimental evidence in object categorization, face recognition [47], language processing and action planning [48] suggests that the different layers of an acquired hierarchy tend to be embedded in spatially segregated circuits. Although a complex hierarchy can be embedded in a local circuit, spatial segregation of different levels of structured abstraction presents key experimental advantages because learning using the discrete hypothesis spaces at each level can be perturbed and observed.

The notion that information in neural circuits may be represented in the form of probability distributions has likewise received attention and theories for how circuit computations can operate on probability distributions have been forwarded [49]. Of the proposed frameworks, Monte Carlo, or stochastic sampling-based approximation is particularly well-suited to support not only inference but also probabilistic learning [50], and can be extended to problems in which learning operates on richly structured abstractions. For a discrete hypothesis space at each

level of the structured abstraction, sampling-based representation would amount to representing an individual or restricted set of hypotheses — a notion consistent with the emerging view that humans and animals may approach many complex settings by evaluating a small number of hypotheses at any given time [48,51]. Indeed, one could begin to correlate ensemble dynamics with different levels of hierarchy that an animal is evaluating with specific hypotheses (symbols) — inferred from a combination of behavioral readouts and computational modeling. If this proves possible in several cognitive domains then the task of investigating the neural basis of inductive inference is more tractable.

How might one identify neural representations of specific symbols in such structured abstractions? This is difficult even if one assumes that the behavioral task provides a means to assess what the encoded structure must be at specific time points, and even if there are stable symbolic variables related to the creation and use of this structure. Such variables are likely to be encoded at the level of neural ensembles rather than single cells (although see Wallis *et al.* for examples of how even single unit recordings can provide windows into what are likely to be more complex population representations of

abstract task parameters [52]). Further, although it is possible that stable variables are encoded in the form of stable neural ensemble activity patterns [53], recent experimental and theoretical work suggests this need not be the case; stable variables can also be represented by dynamically varying patterns of ensemble activity [54•,55,56]. An attempt by McKenzie *et al.* to extract hierarchies of neural representation in the context of transfer learning supports the existence of a hierarchy of neural manifolds, although its relationship to the task structure remains unclear [25•]. Luckily for the field, identifying stable low-dimensional trajectories and manifolds in high-dimensional space using a combination of large-scale recordings and population analyses is becoming increasingly tractable.

One of the key features distinguishing NPHBMs from other approaches is the ability to grow structures non-parametrically as needed to accommodate additional observations. As the complexity of the task increases, the potentially intractable search for the right structure is constrained by a prior on relevant aspects of the structure. There are intriguing hints of such processes in animal behavior and physiology across species.

The prediction that the structured abstraction of the environment should grow nonparametrically has recently been tested in a study of spatial representations in the hippocampus. By exposing rats to a track of increasing lengths, Rich *et al.* (2014) investigated the process by which hippocampal place cells represent environments of different sizes [57••]. Consistent with a process that grows the inferred structure of the environment with task complexity, Rich *et al.* found that the number of neurons recruited was logarithmically related to the track length. This finding supports the view that neural circuits may not only represent acquired structured abstractions of the environment but also flexibly grow them according to task demands.

Recent efforts to use closed-loop brain computer interfaces (BCIs) to study behaviorally relevant neural dynamics have investigated the constraints that neural manifolds place on learning. These studies fit with the thesis that neural manifolds are key components of learned structures. In an intriguing recent study, Sadtler *et al.* (2014) trained non-human primates on a task that required the animals to alter the neuronal dynamics of their primary motor cortices to move a cursor on a screen to one of eight targets [54••]. The closed-loop nature of the task allowed Sadtler *et al.* to specify how neural activity changes mapped to cursor movements. Remarkably, primates could easily learn to change their neuronal dynamics along the intrinsic manifold, but showed difficulty learning to think with activity patterns outside of it. Restriction of network dynamics to specific manifolds could be one of the neural instantiation of structured priors, and BCI

techniques could help expose the specific priors animals might possess.

Although these recent studies provide an encouraging start, they also highlight the challenge of uncovering the neural implementation of sophisticated mental representations of structured knowledge that is not directly accessible to the experimenter. For significant further progress to be made, experimenters must be able to correctly infer the specific structured abstraction an animal is using to solve a task. One solution is to use computational models to infer the structured abstraction from behavior. However, such model-fitting based approaches carry risks, as compelling fits to specific models are often insufficient to ascertain that animals rely on the exact computations to solve the task. In a cautionary tale from abstract structure learning in model organisms, the initial interpretation that pigeons and baboons were capable of learning an abstract ‘same’-‘different’ categorization, was later revised when additional behavioral tests discovered that animals, instead, followed a much simpler rule that tracked the degree of variability in the intensity of displayed images [58].

Thus, developing behavioral frameworks that isolates the structured abstraction an animal represents and uses to solve interesting tasks is a key challenge for the field. One path forward would borrow from the rigorous tradition of psychophysics [59] and combine behavioral modeling with a principled way of varying the details of task design so that an assessment could be made whether the behavioral model properly accounts for resulting variable task performance. Attempts to expose the structure of the learned abstraction and the computations done at each level of the structure by combining response deadline procedures with behavioral error analysis have recently been made in human studies [60]. It seems essential for the field going forward to find ways to translate similar methods to animal models.

Conclusions

Since Edward Thorndike’s tests of cats’ escapology from puzzle boxes, there has been a fascination with, and debate about, if and how animals internalize and use the structure of the world [61]. Contemporary cognitive scientists have recruited NPHBM to generate normative descriptions of cognition that infer this structure, but the project of describing how the mind and the brain learn internal models has remained challenging. Despite this, there is promising recent evidence that some of the computational processes described by NPHBM may be implemented by neural circuits.

Acknowledgements

We thank Brenden Lake for contributing Figure 2, Alla Karpova for significant contributions and Shaul Druckmann, Maksim Manakov, Mikhail Proskurin, Vivek Jayaraman and other colleagues at Janelia Research Campus for comments on and assistance with the manuscript. DGRT is funded by Howard Hughes Medical Institute.

References and recommended reading

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

- of special interest
 - of outstanding interest
1. Gelman A, Hill J: *Data analysis using regression and multilevel/hierarchical models*. Cambridge; New York: Cambridge University Press; 2007.
 2. Chomsky N: *Aspects of the theory of syntax*. Cambridge: MIT Press; 1965.
 3. Jordan MI: *Learning in Graphical Models*. MIT Press; 1998.
 4. Gershman SJ, Blei DM, Niv Y: **Context, learning, and extinction**. *Psychol Rev* 2010, **117**:197-209.
 5. Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND: **How to grow a mind: statistics, structure, and abstraction**. *Science* 2011, **331**:1279-1285.
 6. Braun DA, Mehring C, Wolpert DM: **Structure learning in action**. *Behav Brain Res* 2010, **206**:157-165.
 7. Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C: **The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees**. *Neuron* 2015, **88**:2-19.
 8. Miller EK, Cohen JD: **An integrative theory of prefrontal cortex function**. *Annu Rev Neurosci* 2001, **24**:167-202.
 9. Nienborg H, Roelfsema PR: **Belief states as a framework to explain extra-retinal influences in visual cortex**. *Curr Opin Neurobiol* 2015, **32**:45-52.
 10. Feynman RP, Leighton R, Hutchings E: *Surely you're joking, Mr. Feynman!: adventures of a curious character*. Random House; 1992.
 11. Shettleworth SJ: **Clever animals and killjoy explanations in comparative psychology**. *Trends Cogn Sci* 2010, **14**:477-481.
 12. Dennett D: **Darwin's strange inversion of reasoning**. *Proc Natl Acad Sci U S A* 2009, **106**(Suppl 1):10061-10065.
 13. Pinker S, Jackendoff R: **The faculty of language: what's special about it?** *Cognition* 2005, **95**:201-236.
 14. Buss D: *Evolutionary psychology: the new science of the mind*. Fifth edition. Taylor & Francis; 2015.
 15. Grosenick L, Clement TS, Fernald RD: **Fish can infer social rank by observation alone**. *Nature* 2007, **445**:429-432.
 16. Nieder A: **Coding of abstract quantity by 'number neurons' of the primate brain**. *J Comp Physiol A: Neuroethol Sens Neural Behav Physiol* 2013, **199**:1-16.
 17. Whiten A, Horner V, De Waal FB: **Conformity to cultural norms of tool use in chimpanzees**. *Nature* 2005, **437**:737-740.
 18. Povinelli DJ, Ballew NG: *World without weight: perspectives on an alien mind*. Oxford University Press; 2012.
 19. Seed AM, Call J, Emery NJ, Clayton NS: **Chimpanzees solve the trap problem when the confound of tool-use is removed**. *J Exp Psychol: Anim Behav Process* 2009, **35**:23.
 20. Ditz HM, Nieder A: **Neurons selective to the number of visual items in the corvid songbird endbrain**. *Proc Natl Acad Sci U S A* 2015, **112**:7827-7832.
 21. Cheke LG, Bird CD, Clayton NS: **Tool-use and instrumental learning in the Eurasian jay (*Garrulus glandarius*)**. *Anim Cogn* 2011, **14**:441-455.
 22. Weir AA, Chappell J, Kacelnik A: **Shaping of hooks in New Caledonian crows**. *Science* 2002, **297**:981.
 23. Laurent V, Balleine BW: **Factual and counterfactual action-outcome mappings control choice between goal-directed actions in rats**. *Curr Biol* 2015, **25**:1074-1079.
 24. Tse D, Langston RF, Kakeyama M, Bethus I, Spooner PA, Wood ER, Witter MP, Morris RG: **Schemas and memory consolidation**. *Science* 2007, **316**:76-82.
 25. McKenzie S, Frank AJ, Kinsky NR, Porter B, Riviere PD, Eichenbaum H: **Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas**. *Neuron* 2014, **83**:202-215.
Using a transfer learning paradigm where learning in one context facilitated learning in another, the authors examined hippocampal activity while animals performed the task.
 26. Beckers GJL, Bolhuis JJ, Okanoya K, Berwick RC: **Birdsong neurolinguistics: songbird context-free grammar claim is premature**. *Neuroreport* 2012, **23**:139-145.
 27. Abe K, Watanabe D: **Songbirds possess the spontaneous ability to discriminate syntactic rules**. *Nat Neurosci* 2011, **14**:U1067-U1173.
 28. Harlow HF: **The formation of learning sets**. *Psychol Rev* 1949, **56**:51-65.
 29. Braddick O: **A short-range process in apparent motion**. *Vision Res* 1974, **14**:519-527.
 30. Anderson JR: *The adaptive character of thought*. Psychology Press; 1990.
 31. Kemp C, Perfors A, Tenenbaum JB: **Learning overhypotheses with hierarchical Bayesian models**. *Dev Sci* 2007, **10**:307-321.
 32. Goodman ND, Ullman TD, Tenenbaum JB: **Learning a theory of causality**. *Psychol Rev* 2011, **118**:110-119.
 33. Jones SS, Smith LB: **How children know the relevant properties for generalizing object names**. *Developmental Sci* 2002, **5**:219-232.
 34. Tsivavidis P, Gershman SJ, Tenenbaum JB, Schulz L: **Information selection in noisy environments with large action spaces**. In *Proceedings of the thirty-sixth annual conference of the cognitive science society*. 2014.
 35. Pacer M, Griffiths L: **Elements of a rational framework for continuous-time causal induction**. *Proc. Annu. Conf. Cogn. Sci. Soc.* 2012:833-838.
 36. Ullman TD, Goodman ND, Tenenbaum JB: **Theory learning as stochastic search in the language of thought**. *Cognitive Development* 2012, **27**:455-480.
 37. Gershman SJ, Norman KA, Niv Y: **Discovering latent causes in reinforcement learning**. *Curr Opin Behav Sci* 2015, **5**:43-50.
 38. Lake BM, Salakhutdinov R, Tenenbaum JB: **Human-level concept learning through probabilistic program induction**. *Science* 2015, **350**:1332-1338.
The authors developed a model that learned the stroke-by-stroke casual structure of character generation. The final model captured the ability to generate characters from a single exposure using the composition of strokes.
 39. Griffiths TL, Tenenbaum JB: **Optimal predictions in everyday cognition**. *Psychol Sci* 2006, **17**:767-773.
 40. Gershman SJ, Tenenbaum JB, Jakel F: **Discovering hierarchical motion structure**. *Vision Res* 2015.
 41. McClelland JL, Botvinick MM, Noelle DC, Plaut DC, Rogers TT, Seidenberg MS, Smith LB: **Letting structure emerge: connectionist and dynamical systems approaches to cognition**. *Trends Cognitive Sci* 2010, **14**:348-356.
 42. Bernstein NA: *The co-ordination and regulation of movements*. 1967.
 43. Lashley KS: *The neuropsychology of Lashley; selected papers*. New York: McGraw-Hill; 1960.
 44. Badre D, D'Esposito M: **Is the rostro-caudal axis of the frontal lobe hierarchical?** *Nat Rev Neurosci* 2009, **10**:659-669.
 45. Botvinick MM: **Hierarchical reinforcement learning and decision making**. *Curr Opin Neurobiol* 2012, **22**:956-962.
 46. Huys QJ, Lally N, Faulkner P, Eshel N, Seifritz E, Gershman SJ, Dayan P, Roiser JP: **Interplay of approximate planning strategies**. *Proc Natl Acad Sci U S A* 2015, **112**:3098-3103.

47. Tsao D: **The Macaque Face Patch System: A Window into Object Representation.** *Cold Spring Harb Symp Quant Biol* 2014, **79**:109-114.
48. Koehlin E: **Prefrontal executive function and adaptive behavior in complex environments.** *Curr Opin Neurobiol* 2015, **37**:1-6.
49. Pouget A, Beck JM, Ma WJ, Latham PE: **Probabilistic brains: knowns and unknowns.** *Nat Neurosci* 2013, **16**:1170-1178.
50. Fiser J, Berkes P, Orban G, Lengyel M: **Statistically optimal perception and learning: from behavior to neural representations.** *Trends Cogn Sci* 2010, **14**:119-130.
51. Vul E, Goodman N, Griffiths TL, Tenenbaum JB: **One and done? Optimal decisions from very few samples.** *Cognitive Sci* 2014, **38**:599-637.
52. Wallis JD, Anderson KC, Miller EK: **Single neurons in prefrontal cortex encode abstract rules.** *Nature* 2001, **411**:953-956.
53. Bathellier B, Ushakova L, Rumpel S: **Discrete neocortical dynamics predict behavioral categorization of sounds.** *Neuron* 2012, **76**:435-449.
54. Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-
●● Kabara EC, Yu BM, Batista AP: **Neural constraints on learning.** *Nature* 2014, **512**:423-426.
Using a closed-loop brain computer interface where animals were required to alter the activity in their primary motor cortex to complete a task, the authors were able to show that the non-human primates were only able to learn to vary their neural activity in a limited way.
55. Druckmann S, Chklovskii DB: **Neuronal circuits underlying persistent representations despite time varying activity.** *Curr Biol* 2012, **22**:2095-2103.
56. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV: **Neural population dynamics during reaching.** *Nature* 2012, **487**:51-56.
57. Rich PD, Liaw HP, Lee AK: **Place cells. Large environments ●● reveal the statistical structure governing hippocampal representations.** *Science* 2014, **345**:814-817.
By varying the length of a linear track the authors were able to observe a log-like recruitment of neurons into the hippocampal place field representation.
58. Wasserman EA, Fagot J, Young ME: **Same-different conceptualization by baboons (*Papio papio*): the role of entropy.** *J Comp Psychol* 2001, **115**:42-52.
59. Fetsch CR: **The importance of task design and behavioral control for understanding the neural basis of cognitive functions.** *Curr Opin Neurobiol* 2016, **37**:16-22.
60. Ranti C, Chatham CH, Badre D: **Parallel temporal dynamics in hierarchical cognitive control.** *Cognition* 2015, **142**:205-229.
61. Thorndike EL: *Animal intelligence: Experimental studies.* Macmillan; 1911.
62. Kemp C: *The acquisition of inductive constraints.* Massachusetts Institute of Technology; 2008.