Structured Cognitive Representations and Complex Inference in Neural Systems

Samuel J. Gershman, Joshua B. Tenenbaum ({sjgershm,jbt}@mit.edu)
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA 02139 USA

Alexandre Pouget (Alexandre.Pouget@unige.ch)
Department of Neuroscience, University of Geneva
CH-1211 Geneva 4 Switzerland

Matthew Botvinick (matthewb@princeton.edu)
Department of Psychology, Princeton University
Princeton, NJ 08540 USA

Peter Dayan (dayan@gatsby.ucl.ac.uk)
Gatsby Computational Neuroscience Unit, University College London
London WC1N 3AR United Kingdom

Keywords: Bayesian models, rational analysis, perception, olfaction, memory

Summary

The dream of cognitive neuroscience has always been a seamless integration of cognitive representations with neural machinery, but—despite decades of work—fundamental gaps remain. Part of the problem is that many contemporary theories of cognition are formulated in terms of representations and computations that are quite different from those used in computational neuroscience. Bridging this gap requires more than simply a translation between theoretical concepts in the two fields; what is needed is a more radical updating of neuroscience's theoretical vocabulary.

What should this vocabulary look like? Some important features of representations and computations used in contemporary cognitive theories are:

- Compositional, recursive and relational representations (Fodor, 1975; Smolensky, 1990; Hummel & Holyoak, 2003; Stewart et al., 2011).
- Flexible use of different structural forms (e.g., taxonomic vs. causal knowledge; Kemp & Tenenbaum, 2009).
- Multiple levels of abstraction (Tenenbaum et al., 2011).
- Knowledge partitioning / clustering (Lewandowsky & Kirsner, 2000).
- Complex intuitive theories (e.g., naive physics, theory of mind; Carey, 2009).
- Algorithms that operate on these representations (e.g., dynamic programming, Monte Carlo methods; Griffiths et al., 2012).

These representations and computations are "structured" in the sense that they incorporate rich domain knowledge and strong constraints (Tenenbaum et al., 2011).

This symposium addresses the question: how do neural circuits acquire and compute with structured representations? This question is examined from a number of angles. **Gershman** introduces the basic issues and discusses attempts to articulate a neurally plausible theory of structured cognition. **Pouget** describes recent work on implementing complex probabilistic computations in neural circuits. **Botvinick** shows how neural circuits can be used to discover hierarchical task structure in the environment. Finally, **Dayan** discusses work on wedding richly structured models of semantics with representations of individual episodes. Each talk will be 20 minutes long, followed by a 20 minute panel discussion with speakers moderated by **Tenenbaum**.

Gershman: from knowledge to neurons

How can neurons express the structured knowledge representations central to intelligence? This problem has been attacked many times from various angles. I discuss the history of these attempts and situate our current understanding of the problem. I then outline a new approach based on the idea of *compressing* structured knowledge using neurons in a way that supports probabilistic inference. I illustrate this approach using examples from motion perception and value-based decision making.

Pouget: modeling the neural basis of complex intractable inference

It is becoming increasingly clear that neural computation can be formalized as a form of probabilistic inference. Several hypotheses have emerged regarding the neural basis of these inferences, including one based on a type of code known as probabilistic population codes or PPCs (Ma et al., 2006). PPCs have been used to model simple forms for multisensory integration, attentional search, perceptual decision making or causal inference, for which human subjects have been shown to be nearly optimal. However, most inferences performed by the brain are too complex be solved optimally in a reasonable amount of time and must therefore involve approximate solutions. We have started to explore how neural circuits could implement a particular form of approximation, called variational Bayes, with PPCs (Beck et al., 2012). Remarkably, this approximation requires a nonlinearity known as divisive normalization which has already been found in most neural circuits. This approach can be applied to a wide range of complex inferences, such as the ones involved in olfactory processing, image processing in the primary visual cortex and other related problems.

Botvinick: discovering hierarchical task structure

Naturalistic action displays a hierarchical structure: Simple actions cohere into subtask sequences or component skills, which in turn combine to realize overall goals. Computational models from cognitive psychology, artificial intelligence, and most recently neuroscience, have sought to characterize the representations and mechanisms underlying hierarchical action control (Botvinick, 2008). However, such models tend to neglect a fundamental question: How do hierarchical representations of action or task structure initially arise? We approach this as a learning problem, asking how useful component skills can be inferred from experience. Behavioral evidence suggests that such learning arises from a structural analysis of encountered problems, one that maximizes representational efficiency and, as a direct result, decomposes task into subtasks by 'carving' them at their natural 'joints.' A key question is how this analysis and optimization process might be implemented neurally. Recent data suggests an intriguing answer: Detection of hierarchical task structure might arise as a natural consequence of predictive representation. I'll present computational work fleshing out this possibility, along with behavioral and fMRI data that lend it considerable initial support.

Dayan: unsupervised learning and the representation of episodic structure

The representation of hierarchically structured knowledge in systems using distributed patterns of activity is an abiding concern for the connectionist solution of cognitively rich problems. One particularly important unresolved issue concerns episodic versus semantic structure—how rich generative models of the semantics of domains can be used in the representation of particular, structured, entities. I will unpack this problem and suggest some routes to solutions.

References

Beck, J., Heller, K., & Pouget, A. (2012). Complex inference in neural circuits with probabilistic population codes and topic models. In *Advances in neural information processing systems 25* (pp. 3068–3076).

Botvinick, M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208.

Carey, S. (2009). *The origin of concepts*. Oxford University Press, USA.

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Griffiths, T., Vul, E., & Sanborn, A. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science*, *21*(4), 263–268.

Hummel, J., & Holyoak, K. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*(2), 220–264.

Kemp, C., & Tenenbaum, J. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*(1), 20–58.

Lewandowsky, S., & Kirsner, K. (2000). Knowledge partitioning: Context-dependent use of expertise. *Memory & Cognition*, *28*(2), 295–305.

Ma, W., Beck, J., Latham, P., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, *46*(1), 159–216.

Stewart, T., Bekolay, T., & Eliasmith, C. (2011). Neural representations of compositional structures: representing and manipulating vector spaces with spiking neurons. *Connection Science*, *23*(2), 145–153.

Tenenbaum, J., Kemp, C., Griffiths, T., & Goodman, N. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.