Moving Beyond Where and What to *How*: Using Models and fMRI to Understand Brain-Behavior Relations

Bradley C. Love (brad_love@mail.utexas.edu)
Department of Psychology
Austin, TX 78712 USA

John P. Spencer (john-spencer@uiowa.edu)
Department of Psychology and Delta Center, E11 Seashore Hall
Iowa City, IA 52242 USA

Symposium Overview

One central goal of cognitive science is to understand how the brain supports cognition. Toward this end, a great deal of effort is devoted toward computational modeling and brain imaging. The former effort is well represented at the Annual Conference, whereas the latter effort is neglected. One common criticism of brain imaging research from the cognitive science community is that it is overly focused on the "where" of cognition, as opposed to the "how" (i.e., process-level questions linking brain and behavior). Model-based analysis of fMRI data links models to the interpretation of imaging data, allowing process-level questions to be asked. The basic approach involves fitting models to behavioral data and then using internal quantities from the models as regressors in the imaging analysis. In this symposium, a broad assortment of leading researchers demonstrate the value of this approach in several domains.

Keywords: Computational models, reinforcement learning, categorization, working memory, social cognition, fMRI.

Speakers

We have assembled top speakers with extensive expertise in computational modeling of behavioral and neural data:

Nathaniel Daw, Assistant Professor, New York University. Dr. Daw's research concerns reinforcement learning and decision making from a computational approach, and particularly the application of computational models to the analysis of behavioral and neural data.

Bradley C. Love, Professor of Psychology, University of Texas. Dr. Love is expert in experimental and computational explorations of learning and decision making.

John O'Doherty, Professor, Caltech. His main research focus is on the neural mechanisms underpinning reinforcement-learning and value-based decision making.

John P. Spencer **(Panel Moderator)**, Professor, University of Iowa. Dr. Spencer is an expert in the use of dynamic neural fields to capture behavioral and neural data.

Speaker Abstracts

The abstracts provide broad coverage of topics, including working memory, reinforcement learning, category learning, and social inference. This diversity of problem domains will make clear the commonalities and general applicability of model-based analysis of brain data. At the same time, the particularly challenges faced in each domain will be informative and raise discussion topics.

Testing a dynamic neural field model of visual working memory with fMRI (Spencer, Buss & Magnotta)

Efficient visually-guided behavior depends on our ability to form, retain, and compare visual representations that may be separated in space and time. This ability relies on visual working memory (VWM). Although research has begun to shed light on the neuro-cognitive systems subserving this form of memory, few theories have addressed these processes in a neurally-grounded framework.

Here, we describe a layered neural architecture that captures the cortical population dynamics that underlie VWM, including the encoding, maintenance and comparisons operations involved in change detection. We then test this model using functional neuroimaging. Recent work has shown that the BOLD response is strongly correlated with local field potentials (LFPs). An analog of LFPs can be estimated from dynamic neural field models. This estimate can be convolved with an impulse response function to yield time-dependent hemodynamic predictions.

Using this approach, we show that the DFN model quantitatively captures fMRI data from recent studies probing changes in the BOLD response in the intraparietal sulcus (IPS) as set size increases in change detection, as well as data showing stronger activation on change trials versus same trials. We also test a novel prediction of the model that BOLD responses should be greater on false alarms versus misses. These data run counter to common explanations of the origin of errors in change detection.

Computational models as neural hypotheses: Reinforcement learning (Daw)

The predominant methods for analyzing neuroimaging data center on assessing explicit statistical models of the neural response. I consider how this approach can be extended to test more psychological or functional level models of neural computation. The function I focus on is learned trial-and-error decision making. Computational algorithms for this function – known in computer science as reinforcement learning -- can be viewed as explicit hypotheses about how

subject behavior and associated neural responses (e.g., BOLD signals related to reward predictions or prediction errors) may change, trial by trial, with feedback. These hypotheses can be tested and refined using standard model comparison and parameter estimation techniques. I first discuss, methodologically, how to frame these tests in the context of fMRI analysis, dealing with problems such as model selection in the random effects setting and free parameters that affect the data nonlinearly. Second, I present recent results from our laboratory in which we use these techniques to study the trial-by-trial time course of learning in computationally challenging decision tasks. In particular, we consider how and whether different types of information -- about experienced and foregone rewards, their variance and covariance, and sequential task structure -- differentially impact choice behavior and BOLD signals in decision-related areas such as striatum.

Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members (Love, Davis, & Preston)

Formal models have proven critical in understanding the cognitive psychology of category learning. Here, we use these cognitive models to advance the cognitive neuroscience of category learning.

Category knowledge can be explicit, yet not conform to a perfect rule. For example, a child may acquire the rule "If it has wings, then it is a bird," but then must account for exceptions to this rule, such as bats. The current study explored the neurobiological basis of rule-plus-exception learning by using quantitative predictions from a category learning model, SUSTAIN, to analyze behavioral and functional magnetic resonance imaging (fMRI) data. SUSTAIN predicts that exceptions require formation of specialized representations to distinguish exceptions from rule-following items in memory. By incorporating quantitative trial-by-trial predictions from SUSTAIN directly into fMRI analyses, we observed medial temporal lobe (MTL) activation consistent with two predicted psychological processes that enable exception learning: item recognition and error correction. SUSTAIN explains how these processes vary in the MTL across learning trials as category knowledge is acquired. Importantly, MTL engagement during exception learning was not captured by an alternate, exemplar-based model of category learning, or by standard contrasts comparing exception and rule-following items. The current findings thus provide a well-specified theory for the role of the MTL in category learning where the MTL plays an important role in forming specialized category representations appropriate for the learning context.

Computational model-based fMRI of social inference and learning (O'Doherty)

In model-based functional magnetic resonance imaging (fMRI), signals derived from a computational model for a specific cognitive process are correlated against fMRI data from subjects performing a relevant task to determine brain regions showing a response profile consistent with that model. In this talk I will illustrate the merits of this approach in the light of recent studies in the domain of social cognition.

A fundamental capacity underlying much of human social processing is the ability to "mentalize" or infer the thoughts or intentions of others. Human neuroimaging studies have shown that specific brain structures are engaged during mentalizing such as the dorsomedial prefrontal cortex and posterior superior temporal sulcus. However, very little is known about the putative computational processes being implemented in these regions in order to underpin such a capacity. Here I will demonstrate how the application of a formal computational model capable of learning to make predictions based on the mental states (or beliefs) of others can when combined with neuroimaging data, reveal specific computational roles for each component of the mentalizing network. I will further review evidence for the existence of computational signals in the brain capable of mediating learning about the value of stimuli in the world through observation of the experiences of others. Collectively these studies illustrate how model-based fMRI can potentially provide insights into how a particular cognitive process is implemented in a specific brain area as opposed to merely identifying where a particular process is located.

Acknowledgments

This research was made possible by NIH MH062480 and NSF BCS-1029082 awarded to JPS; NIMH 1R01MH087882-01, part of the CRCNS program, and a McKnight Scholar Award to NDD, AFOSR Gant #FA9550-10-1-0268to BCL, and Gordon and Betty Moore Foundation, Science Foundation Ireland, NSF grants JPO.

Relevant References

Davis, T., Love, B.C., & Preston, A.R. (in press). Learning the Exception to the Rule: Model-Based fMRI Reveals Specialized Representations for Surprising Category Members. *Cerebral Cortex.*

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011) "Model-based influences on humans' choices and striatal prediction errors." *Neuron*, 69, 1204-1215.

Glaescher, J., & O'Doherty, J.P. (2010). Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews: Cognitive Science,* 1, 501-510.

Johnson, J.S., Spencer, J.P., Luck, S.J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science, 20*, 568-577.