Grow your own representations: Computational constructivism

Joseph L. Austerweil (Joseph.Austerweil@gmail.com) Robert L. Goldstone (rgoldsto@indiana.edu)
Thomas L. Griffiths (Tom_Griffiths@berkeley.edu) Todd Gureckis (todd.gureckis@nyu.edu)
Kevin Canini (kevin@eecs.berkeley.edu) Matt Jones (mcj@colorado.edu)

Keywords: representational change, Bayesian modeling, Connectionism, features, categories

From a cognitivist standpoint, one main interest of psychology is the study of representations of the human mind as they mediate how people react to stimuli in their environment (Palmer, 1978). This can explain why two people that encounter the same stimulus can behave in very different ways (Chomsky, 1959). For example, an art historian viewing a Jackson Pollock painting may exclaim "this is beautiful" due to her representation of his work as a rejection of painting with a brush; however, a lay person may say "this is ugly" due to his representation of the painting as a cluttered mess of colors. Without knowledge of the representations of each person in this example, it would be nearly impossible to explain their behavior when interacting with the Jackson Pollock painting.

Over the last three decades, cognitive psychologists have demonstrated that the representations people use can change flexibly to capture changes in their environment (Hoffman & Richards, 1985; Schyns, Goldstone, & Thilbaut, 1998; Goldstone, 2003). However, if the representations we use are determined by the stimuli in our environment, this threatens the explanatory utility of representations as it could be superfluous to use representations to explain people's reaction to stimuli if the representations are determined by the stimuli. Thus, cognitive psychologists need to explicitly formulate how representations change with experience.

Although computational modelers, from connectionists to Bayesians, disagree on many things, one thing they do agree on is the importance of representations in their models (McClelland et al., 2010; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). Recently, there has been a growing interest in exploring computational models that adapt their representations with experience in ways that match this human capacity. In this symposium, we explore computational models that adapt their representations with experience in ways that are inspired by the human capability.

Recently, there have been several proposals for computational models whose representations flexibly adapt to the input data like people do; however, there has not been a thorough comparison of the different models. The goal of the symposium is the compare and contrast the different methods, evaluate their ability to capture of human representation learning, and make explicit what is meant in each model by "representation change" as this can be a controversial claim (Schyns et al., 1998). Currently, it is not clear whether or not the different proposals mean the same thing by a "representation" and if they are competing proposals to explain the same aspect of human cognition or different levels of explanation. Thus, the symposium will emphasize understanding what is meant by representation change and how well each model can explain human representation change.

The symposium will focus on a wide variety of methods for representation learning from some of the most popular computational paradigms in computational cognitive science: nonparametric Bayesian modeling (Austerweil & Griffiths; Canini & Griffiths), connectionist modeling (Gureckis; Goldstone), and reinforcement learning (Jones). Importantly, each presenter will focus on how their computational proposals explain human experimental data and discussing what exactly is a representation in their framework and how they are inferred. Thus, the symposium should be interesting to a broad audience of cognitive scientists (from computation modelers to experimentalists to philosphers). We hope it inspires a growth of new computational models and human experiments in this underdeveloped, yet incredibly important, aspect of cognitive science.

Introduction and Nonparametric Bayesian Models of Feature Learning

Austerweil and Griffiths Cognitive psychology is concerned primarily with representations and how they mediate the response to stimuli. In this talk, we present a framework for exploring the principles underlying human feature learning using nonparametric Bayesian statistics. We show that our framework can capture how people infer features using statistical information of the observed images, spatial information from the observed images, and categorization cues. Next, we extend our initial framework to infer features that are invariant over a set of transformations and demonstrate that the model infers new invariant features like people do. Although most shapes and features can be transformed by translations and rescalings, some shapes and features lose their identity when rotated. We show how our model is easily extended to capture how people infer the allowable set of transformations of an object from their observations of the object. Finally, we conclude with the implications of our framework for reference frames in shape perception and feature-based cognitive models and compare it to other approaches for inferring representations.

Building flexible categorization models by grounding them in perception

Goldstone One limitation of most existing models of categorization is that they do not start with a perceptually grounded representation of the objects that they categorize. Instead, they use dimensional or featural representations that discard information about the spatial relations among an object's parts. This restricts the models' ability to create psychologically plausible object representations that can be flexibly adapted to meet categorization demands. I will describe a neural network model, C-PLUS, that creates part-based representations of objects that honor perceptual constraints such as proximity and good continuation. Using a modified competitive learning algorithm for object segmentation, it decomposes a set of incrementally presented objects into parts that can be composed together to regenerate the set of objects to be categorized. These parts are learned at the same time that weights from the parts to categories are learned, allowing perceptual representations not only to guide categorization, but categorization to guide perceptual representations as well. The model is applied experimental results on the unitization of object elements into complex wholes, learned differentiation of originally fused encodings into parts, and experience-dependent changes to selective attention abilities.

Constructing representations through reinforcement learning by improving generalization

Jones One critical role of representations in cognition is that they determine patterns of similarity, and hence generalization, among stimuli or situations. To the extent that two stimuli have similar representations, past experience with one will have a large influence on the learner's response to the other. Thus a reasonable goal is to develop representations that induce appropriate generalization, in that stimuli with similar consequences or appropriate actions will tend to have similar representations. This connection suggests a mechanism for representation learning, based on improving generalization in response to prediction error. We present a formal framework instantiating this idea, in which representation learning is driven by the temporal-difference (TD) error from reinforcement learning. The model explains patterns of human learning to shift attention among stimulus features, according to how well different features capture the structure of a task. We will also present evidence supporting a counterintuitive prediction of the model in which reduced training can lead to improved asymptotic performance, resulting from order effects that emerge from the model's incremental learning mechanism. This finding illustrates an important advantage of mechanistic modeling over computational-level (e.g., Bayesian) approaches

A nonparametric hierarchical Bayesian framework for modeling human categorization

Canini and Griffiths

Traditional models of human categorization typically fall into one of two groups: prototype models, which use minimally complex category representations, and exemplar models, with maximal complexity. Previous work showed that these can both be described in the framework of probability density estimation. Within this framework, we can identify a new class of psychological models using mixture distributions. Indeed, several researchers have begun to explore this possibility. We present a unifying model for categorization models based on the statistical tool of nonparametric hierarchical Bayesian modeling. The overarching model, called the hierarchical Dirichlet process (HDP), provides a flexible, formal account of category learning for both individual categories and interrelated systems of categories. Its behavior can replicate that of prototype models, exemplar models, and more recent mixture models, as it adjusts the complexity of its representations in response to the observed data. The HDP can also be used to introduce dependencies in the learning of multiple categories, allowing us to give a formal account for previously unexplored aspects of human category learning such as transfer learning and taxonomy induction.

Endnote: Breaking Sticks or Building Clusters? Representation Building, Learning, and the Brain

Gureckis Traditional models of human learning tend to focus on parameter inference, in that learning involves adjusting the internal parameters of an a-priori fixed architecture. However, a key feature of human learning is the discovery and growth of new representations that help us to interpret and interact with the world. The work reviewed in this symposium offers at least two distinct ways of thinking about this psychological process. Innovations in non-parametric Bayesian statistics have ushered a new generation of probabilistic models that can flexibly adjust the complexity of their representation using stochastic process priors (e.g., the "stick breaking process"). Other theorists take a bottom-up approach to representation building, focusing on the incremental learning mechanisms that give rise to representational change (e.g., incremental clustering models). In my talk, I explore the tension between these two approaches using examples from the categorization and sequential pattern learning literatures. I place a particular emphasis on the psychological content of each approach as well as consistency with the neural systems thought to be involved in particular types of representation building (e.g., episodic memory systems). Ultimately, I argue that the gulf between these approaches need not be wide, if both sets of theorists are clearer about the critical importance of the inference mechanism used to drive predictions in their models.

References

Chomsky, N. (1959). A review of B.F Skinner's Verbal Behavior. *Language*, *31*, 26-58.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In *Perceptual Organization in Vision: Behavioral and Neural Perspectives* (p. 233-278). Mahwah, NJ: Lawerence Erlbaum Associates.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357-364.

Hoffman, D. D., & Richards, W. A. (1985). Parts in recognition. *Cognition*, *18*, 65-96.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*(8), 348-356.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In *Cognition and categorization* (p. 250-303). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schyns, P., Goldstone, R. L., & Thilbaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1-54.