Rumelhart Symposium: Integrating Human and Machine Vision: In Honor of Shimon Ullman

Symposium Organizer:
Michael J. Tarr (Michael_Tarr@brown.edu)

Symposium Presenters:
Marlene Behrmann (behrmann@andrew.cmu.edu)
Center for the Neural Basis of Cognition
Department of Psychology, Carnegie Mellon University
Pittsburgh, PA 15213 USA

Christof Koch (koch@klab.caltech.edu)
Lois and Victor Troendle Professor of Cognitive and Behavioral Biology
Computation and Neural Systems,
California Institute of Technology
Pasadena, CA 91125 USA

Nikos Logothetis (nikos.logothetis@tuebingen.mpg.de)
Director
Max Planck Institute for Biological Cybernetics
72076 Tübingen, GERMANY

Rafi Malach (Bnmalach@wisemail.weizmann.ac.il)
Morris and Barbara Levinson Professor of Brain Research
Department of Neurobiology,
Weizmann Institute of Science
Rehovot 76100 ISRAEL

Michael J. Tarr (Michael_Tarr@brown.edu)
Fox Professor of Ophthalmology and Visual Sciences
Co-Director, *Center for Vision Research*
Department of Cognitive and Linguistic Sciences,
Brown University
Providence, RI 02912 USA

Keywords: Vision; Object Processing; Computational Neuroscience; Shimon Ullman; Rumelhart Prize.

Introduction

The promise of cognitive science is rooted in combining computational and empirical methods. Shimon Ullman's study of both artificial and biological vision demonstrates the power of this approach in studying the mind and brain. Among his many contributions, Shimon's computational models have provided working systems that offer theoretical accounts of how humans recognize objects, perceive motion, probe their visual world for task-relevant information, and create coherent representations of their environments. At the same time, consideration of the neural bases of primate vision and human visual behavior has provided Shimon with inspiration for his computational models, leading to solutions to difficult problems in artificial intelligence. By learning from biological systems, Shimon has created novel and influential artificial intelligence systems.

Shimon is a fitting recipient of the David E. Rumelhart prize in that his research addresses the theoretical foundations of perception, and draws heavily on both mathematical and experimental investigations, as did the research of David Rumelhart.

Overview of Presentations

Marlene Behrmann

The functional organization of the occipito-temporal visual pathway remains highly controversial. Studies of individuals with impairments in face and/or object recognition present a unique opportunity to understand the necessary and sufficient function of the different ventral regions. Behavioral and imaging (structural, functional and diffusion tensor) data from individuals with acquired and with congenital prosopagnosia will be presented as well as recent findings from the domain of autism. Taken together, the behavioral studies demonstrate that all of these individuals are disproportionately impaired at face processing. However, in all cases, the deficit is more extensive, implicating the perception of other classes of visual objects too, when the demands of the testing are equated across all object classes. Furthermore, a pronounced difficulty in deriving configural (second-order relations) relationships between components of the input may explain the disproportionate difficulty with faces, relative to the other classes of objects. The imaging studies map out an underlying neural circuit, including core regions such as the fusiform gyrus, as well as more distal regions, such as anterior temporal lobe and even parts of frontal cortex, whose integrity is critical for computing the fine-grained configural information required to support normal face processing. The disruption in the integrity of this neural circuit may account for the face and object processing deficits in these impaired individuals.

Christof Koch

What governs the trajectory of covert and overt eye movements? I will describe our system level approach to attentional selection in natural scenes. We implement a bottom-up spatial attentional selection mechanism via a saliency-map representation as originally conceived by Koch and Ullman (1985). Approximately half of the inter-subject variability in eye movement fixations of natural scenes under free viewing conditions can be explained by such a bottom-up approach in static and dynamic scenes. If scenes contain faces, a saliency-based approach with multi-scale, face-specific channels can even be used to decode which subject looked at what scenes based on the subject's eye movements. Finally, we describe how attentional selection can be coupled to a biological-based hierarchical object classification scheme. Current evidence suggests that such a saliency-map is instantiated in the early (< 150 ms) responses of neurons in the lateral intraparietal sulcus and in the frontal eye fields.

Nikos Logothetis

Research in my laboratory concentrates on the neural mechanisms of perception and object recognition. Although our basic research revolves around vision, we are also investigating the relationship between neural activity and perception using other sensory modalities. Such scientific questions require a multimodal methodological approach that integrates information obtained from single units with that derived from mass action potentials as well as from a number of activity-related, surrogate signals such as those monitored during noninvasive neuroimaging. Parallel to our ongoing neuroscientific research, therefore, we are working to develop methodologies that will permit us to the study neural networks in the context of behavioral paradigms.

Rafi Malach

With the advent of fMRI a large body of research has focused on the fundamental issue of the neuronal representation of visual objects. These findings provide important constraints on the type of representation that can be considered. First, each object representation appears to involve a massive activation pattern, likely reaching millions of active neurons for each object image. Second, in high-order object areas, one finds strikingly "holistic" aspects of the representation, such as completion effects and gestalt properties - pointing to neurons optimally responsive to fragments as proposed by Ullman's theoretical work, or even entire object templates. Finally, adaptation experiments reveal an exquisite selectivity of the neuronal responses, even at sub-exemplar levels. Reconciling these different lines of evidence leads to a model in which each neuron in high-order object areas is optimally responsive to a massive number of *different* object templates. The receptive field of such neurons can be metaphorically envisioned as a "totem-pole" where different object templates are stacked on top of each other. Thus, such neurons respond when stimulated with any exemplar in the array of templates they are sensitive to. It is hypothesized that the ambiguity inherent in such multi-template responses is resolved through lateral interactions among the many neurons optimally sensitive to the same object exemplar within the representation.

Michael J. Tarr

How is task-dependent feature diagnosticity reflected in the neural representation of faces? We are investigating these neural codes using the approach to face detection developed by Ullman, Vidal-Naquet and Sali (2002) and extending this method to the task of face individuation. The diagnosticity of facial features (image fragments) for individuation was computed by means of the mutual information between face identity and fragment presence across a set of faces varying in pose and expression. fMRI was then used to explore whether diagnosticity is reflected in the neural responses of face-selective regions. Standard ROI-based contrasts reveal that face-selective regions are sensitive to detection, but not individuation, diagnosticity. In contrast, a pattern classification analysis reveals that bilateral MFG is sensitive to individuation, but not detection, diagnosticity. Thus, neural responses associated with face detection and individuation dissociate between diagnosticity for the two tasks. We speculate that this dissociation reflects differences in task-related computational demands: while detection likely recruits face-selective regions globally, individuation may recruit only subsets within the total neural population of face-selective regions. Such results reinforce the idea that feature codes for visual recognition are computed in a task-specific manner and suggest that image fragments provide a functionally meaningful descriptor of the representations used by the primate visual system.

Acknowledgements

We are grateful to the Robert J. Glushko and Pamela Samuelson Foundation for their support of the Rumelhart Prize and this symposium.

References

Koch C., & Ullman S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219-227.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682-687.