Statistical Learning in Language, Vision, and Comparative Cognition: What have we learned, where are we going?

Elissa L. Newport (newport@bcs.rochester.edu)
Department of Brain & Cognitive Sciences
University of Rochester, Rochester NY 14627 USA

Richard N. Aslin (aslin@cvs.rochester.edu)
Department of Brain & Cognitive Sciences
University of Rochester, Rochester NY 14627 USA

Daniel J. Weiss (djw21@psu.edu)
Department of Psychology, Pennsylvania State University
643 Moore Bldg, University Park PA 16802 USA

Discussant: **Lila Gleitman (gleitman@upenn.edu)**
Institute for Research in Cognitive Science
University of Pennsylvania, Philadelphia PA 19104 USA

Motivation

Statistical approaches to learning have become widespread in the field of language acquisition, and have also been of great interest in studies of high-level vision and of comparative cognition. This symposium will present recent work in each of these fields, with the aim of addressing the more general questions of what we have learned through these lines of research, what questions remain to be answered, and what these approaches can and cannot account for in natural cognition.

Elissa L. Newport

In recent years there has been a major impact on the field of language acquisition from statistically oriented modeling and computational linguistic approaches, as well as from early findings on statistical word segmentation capabilities in human infants. A wide variety of studies have shown that infants, young children, and adults can successfully utilize the statistics of distributional linguistic information not only to find candidate words in a speech stream, but also to form and alter phonetic categories, discover grammatical categories, and acquire simple syntactic structure in miniature languages in the laboratory. At the same time, there have been important questions raised about whether such mechanisms can succeed in real language acquisition, given the complexity of actual linguistic input, and whether such mechanisms must ultimately be combined with qualitatively different, more traditional mechanisms that form symbolic rules or set linguistic parameters. This paper will address these questions by presenting findings from recent studies of statistical learning of syntax, examining the effects of complex multiple cues and also comparing child and adult learners given inconsistent input and (sometimes) forming rule-like generalizations. The results of these studies suggest that a suitably sophisticated statistical learning mechanism can accomplish some of these important aspects of language acquisition.

Richard N. Aslin

Although statistical learning originated within the domain of language learning, questions about its domain-specificity were immediately addressed in studies of tone-sequence learning, which confirmed that it was not restricted to speech stimuli. The present paper presents recent findings on a related question: Is statistical learning modality-specific or does it also apply to visual sequences and visual scenes. Recent studies confirm that the answer is clearly 'yes' and raises the prospect that statistical learning is a general-purpose unsupervised learning mechanism. However, such a mechanism must be highly constrained to enable the extraction of just those statistics that optimally solve a variety of learning tasks. These constraints will be reviewed in studies of both auditory and visual statistical learning in adults and infants. Key constraints include Gestalt principles, social/attentional cues, preferred units of analysis, redundancy reduction in long-term memory, and primacy in the statistics that are extracted from changing structures. This paper will conclude with a discussion of how these constraints are implemented across different domains, modalities, and developmental periods.

Daniel J. Weiss

One of the longstanding debates in the study of language is whether language acquisition requires specialized mechanisms for speech perception or whether it can proceed with the support of generalized auditory mechanisms. In efforts to resolve this issue, over the last 25 years many researchers have adopted a comparative approach, comparing the performance of nonhuman primates and humans on a variety of speech perception tasks thought to be fundamental for language development. Recent findings from this area have reinvigorated the debate of the origins of language by demonstrating that a spate of generalized learning mechanisms, such as the statistical learning mechanism (Hauser, Newport, & Aslin, 2001), appear to be shared across species, as well as providing limited evidence that some computational abilities are exclusively human (e.g., Fitch and Hauser, 2004). This paper will review some of the major findings in this area, and present new data that addresses some of the subtle cross-species differences that permeate even the shared abilities (e.g., Newport, Hauser, Spaepen, & Aslin, 2004). The paper concludes with a discussion of the need for more fine-grained analyses in comparative research, emphasizing the impact of this issue in interpreting behavioral data.

Discussant: Lila Gleitman