Modeling Process Differences in Implicit and Explicit Category Learning: A Symbolic-Connectionist Approach

Leonidas A.A. Doumas
Indiana University

Robert G. Morrison
Northwestern University

Abstract

Much experimental evidence suggests both implicit (i.e., similarity- or feature-based) and explicit (i.e., rule- or relation-based) mechanisms in category learning. Interestingly, these systems may compete and/or interact under certain circumstances. However, most computational models of category learning fail to capture this interaction. We present DORA (Discovery of Relations by Analogy) which accounts for both implicit and explicit learning from a common category learning task (Maddox, Ashby, & Bohill, 2003). We also successfully simulate differential working memory effects on implicit and explicit versions of this task as reported by Zeithamova & Maddox (2006).

Background

- Behavioral (see Ashby & Maddox, 2005); neuropsychological (see Kiri, 2001), and neuroimaging (e.g., Nomura et al., 2007) evidence has suggested that categories can often be learned either via a explicit rule-based mechanism critically dependent on medial temporal and prefrontal brain regions, or via an implicit mechanism relying on the basal ganglia.
- Ashby and Maddox have described one way to construct categorizations tasks based on this dissociation which they characterize as either Rule-Based (RB) or Information Integration (II) based analysis of individual subject response patterns using a Decision-Bound mathematical model.
- Using a paradigm developed by Maddox, Ashby, & Bohill (2003), Zeithamova & Maddox (2006) showed that category structure used to be categorized by rule-based processes showed more sensitivity to distraction by a Stroop working memory distractor task than an implicit version of the task.

Abstract

- DORA (Discovery of Relations by Analogy) is a theory of how we learn relational (i.e., structured) representations from unstructured input.
- Starting with representations of objects attached to features, DORA learns structured representations of single-place predicates and multi-place relations through a process of comparison-based interaction discovery.
- Accounts for over 15 phenomena from the literature on children’s and adults’ relation learning and relational reasoning.

DORA Category Learning

- Implicit - Categorizing
- Explicit - Rule Learning

DORA Simulations

- DORA begins with representations of the items represented as objects attached to features.
- Compares items from same category and learns representations of their common properties (i.e., representations of category defining features
- Uses these representations during categorization
- During categorization, the target item is placed in driver
- Runs either implicit or explicit process

Discussion

- We have presented a first attempt at modeling differences in explicit and implicit category learning using DORA, a model initially constructed to explain how structured representations can be built from unstructured input.
- We have demonstrated that DORA can successfully learn these different category structures using both implicit and explicit mechanisms and that the DORA’s implicit and explicit categorization algorithms are differentially sensitive to changes in working memory.
- DORA provides an opportunity to investigate differences in implicit and explicit categorization mechanisms using a common underlying knowledge representation.

References


The presenters wish to thank John Hummel for helpful discussions and the Indiana University Developmental Training Grant from the NICMH (LAAD), the Northwestern University Mechanisms of Aging and Dementia Training Grant from the NIA (2T32AG020506, RGM), and the Office of Naval Research (SBIR 03-DH09, RGM) for their generous support.