Doumas, L.A.A., & Morrison, R.G. (2007, November). Modeling process differences in implicit and explicit category learning: A symbolic-connectionist approach. Poster presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, California.

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

Leonidas A.A. Doumas
Indiana University

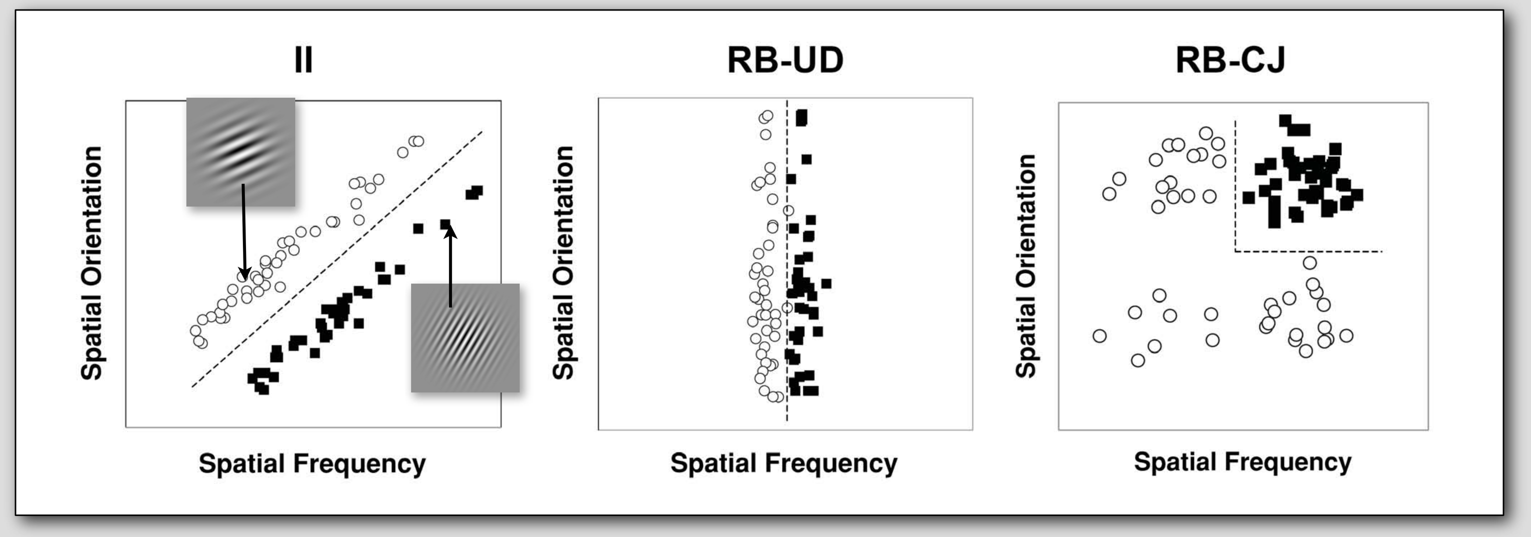
Robert G. Morrison
Northwestern University

Abstract

Much experimental evidence suggests both implicit (i.e., similarity- or feature-based) and explicit (i.e., rule- or relation-based) mechanisms for learning categories. Interestingly, these systems may compete and/or interact under certain circumstances. However, most computational models of category learning fail to capture this interaction. This is because models of implicit and explicit category learning rely on different kinds of representations. Most models of implicit category learning are based on connectionist architectures consisting of networks of distributed units, while, models of explicit category learning typically employ symbolic architectures that capture categories via rules. We present a model of category learning, DORA (Discovery of Relations by Analogy) that uses different learning mechanisms on a single representational substrate. DORA 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 Kéri, 2003), 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 Theory mathematical model.
- Using a paradigm developed by Maddox, Ashby, & Bohill (2003), Zeithamova & Maddox (2006) showed that category structures expected to be categorized by rule-based processes showed more sensitivity to distraction by a Stroop working memory distractor task than an Information Integration category distribution (see plot in figure on 3rd panel to right).



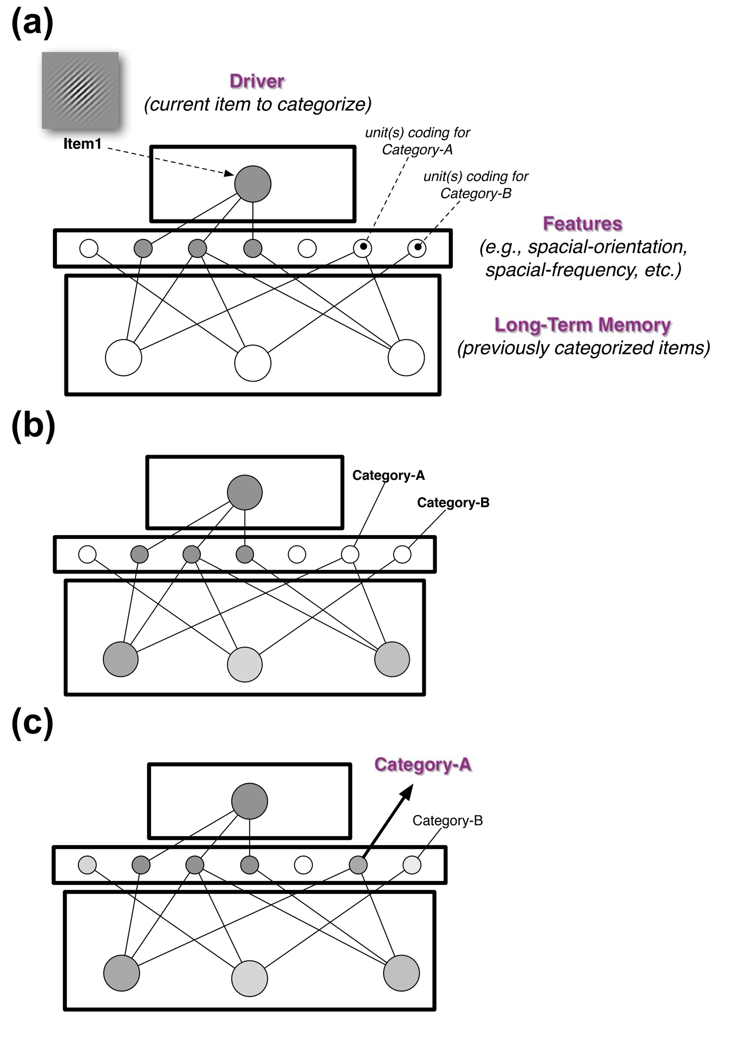
Above - Category structures used by Zeithamova & Maddox (2006)

DORA

- DORA (*Discovery of Relations by Analogy*; Doumas et al., in press) 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 intersection discovery
- Accounts for over 15 phenomena from the literature on children's and adults' relation learning and relational reasoning

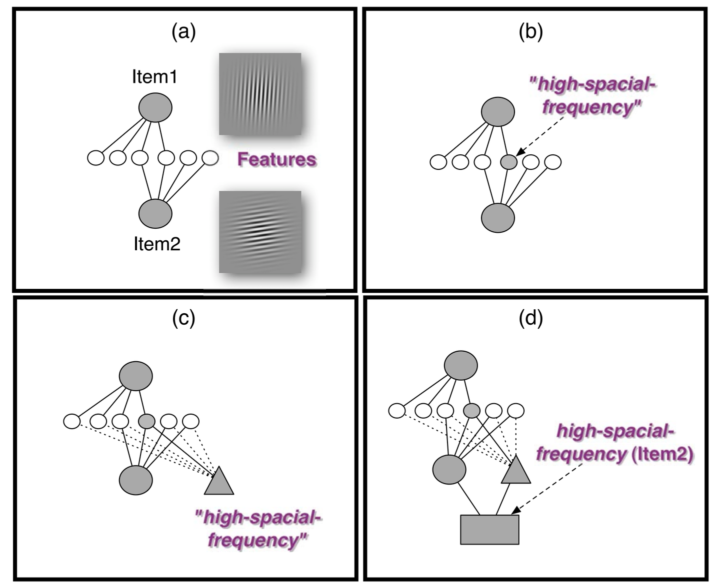
DORA Category Learning

Implicit - Categorizing

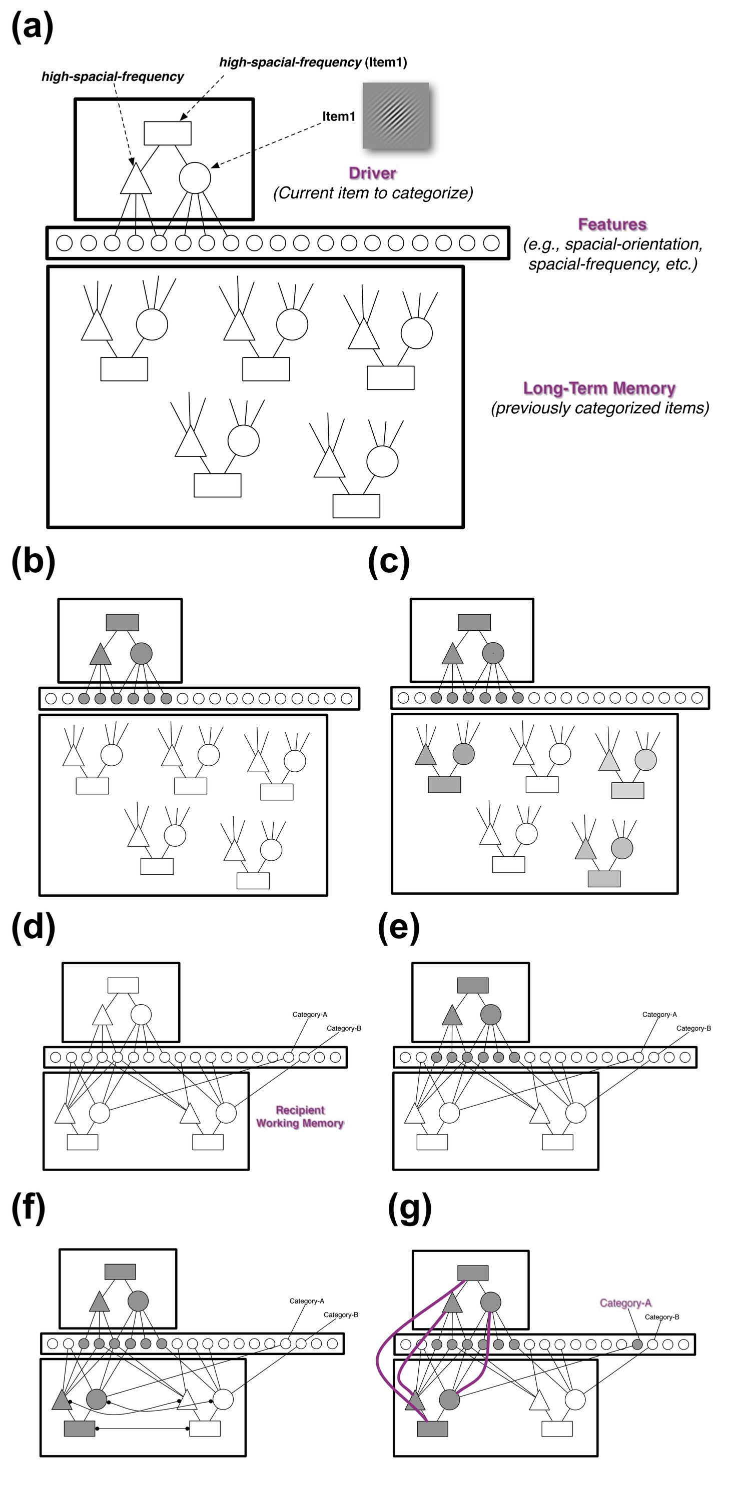


Above - *DORA Implicit Category Learning.* (a) A unit representing the current item to be categorized shares activation with featural units representing its visual characteristics including spatial frequency and spatial orientation. (b) These featural units send activation to other featurally similar items in long-term memory (LTM). (c) Featural units representing category A and B become active depending on whether more category A or B objects are activated in LTM.
Above Right - DORA explicitly predicates a rule, such as "category A is associated with items with higher-spatial frequency" through a process of iterative comparison of items in working memory.
Right - DORA can make categorization decisions based on mapping the current item to past categorized items in LTM. The comparison process depends on precisely firing items in working memory.

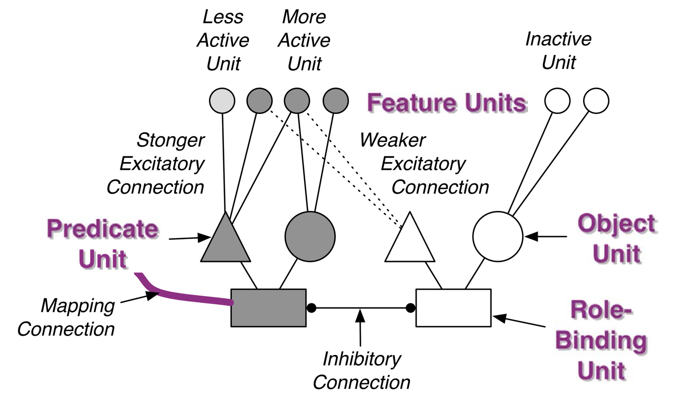
Explicit - Rule Learning



Explicit - Categorizing

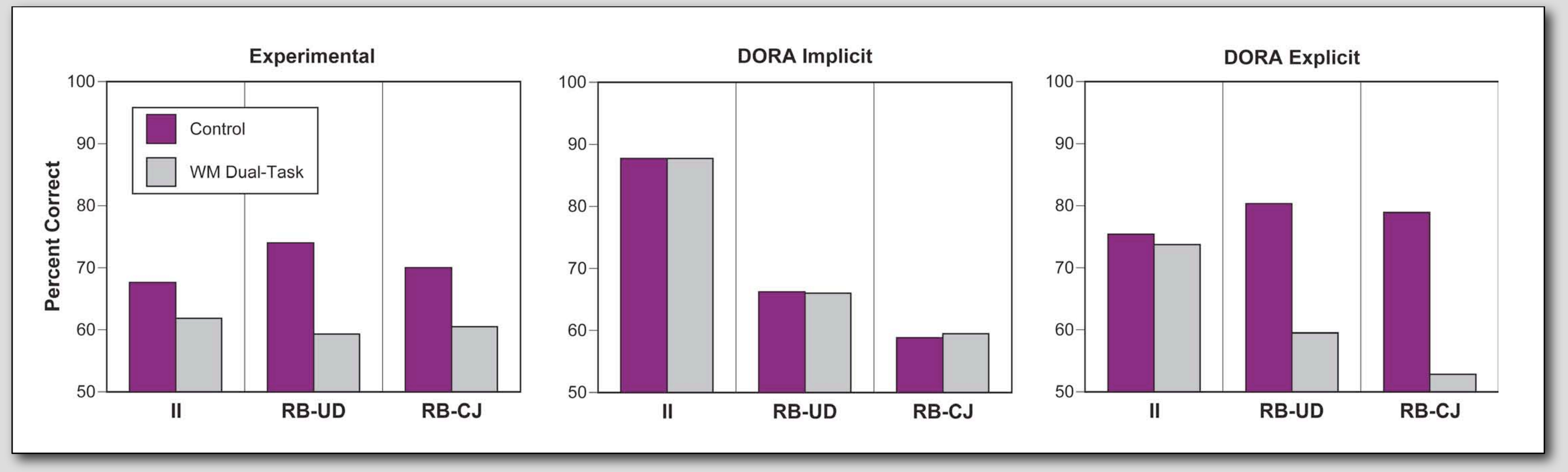


DORA - Legend



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
- Run either implicit or explicit process
- Winner is chosen by feature activation
- To simulate feedback, the target unit is connected to the feature unit representing category-A, if the item was from category-A, or category-B otherwise
- WM load is simulated by reducing lateral inhibition in WM



Above - Experimental data from Zeithamova & Maddox (2006).

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 three different category structures using both implicit and explicit mechanisms, and that the DORA's implicit and explicit categorizing 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

Ashby, F.G., Maddox, W.T. (2005) Human category learning. *Annual Review of Psychology, 56*, 149-78.
Doumas, L.A.A., Hummel, J.E., & Sandhofer, C.M. (in press). A theory of the discovery and predication of relational concepts. *Psychological Review.*
Kéri, S. (2003). The cognitive neuroscience of category learning. *Brain Research Reviews, 43*, 85-109
Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 650-662.
Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., & Parrish, T. B. et al. (2007). Neural correlates of rule-based and information-integration visual category learning. Cerebral Cortex, 17, 37-43.
Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. Memory & cognition, 34, 387-398.

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 OSD03-DH07, RGM) for their generous support.