Linear separability and human category learning: Revisiting a classic study

Kimery Levering[1], Nolan Conaway[2] and Kenneth J. Kurtz[3]

[1]Department of Psychology, Marist College

[2]Department of Psychology, University of Wisconsin, Madison

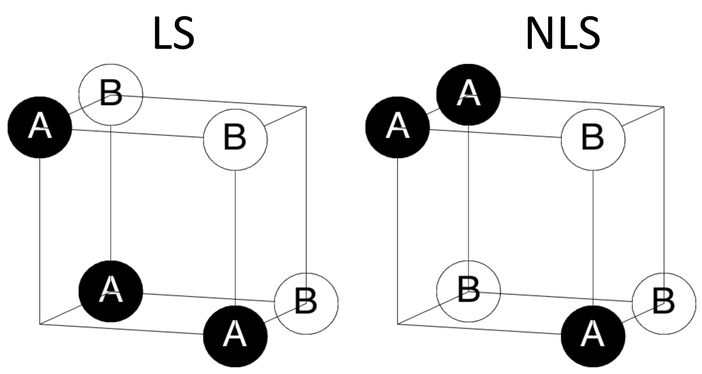
[3]Department of Psychology, Binghamton University

The ability to acquire non-linearly separable (NLS) classifications is well documented in the study of human category learning. In particular, one experiment (Medin & Schwanenflugel, 1981; E4) is viewed as the canonical demonstration that, when within- and between- category similarities are evenly matched, NLS classifications are not more difficult to acquire than linearly separable ones. The results of this study are somewhat at issue due to non-standard methodology and small sample size. We present a replication and extension of this classic experiment. We did not find any evidence of an advantage for linearly separable classifications. In fact, the marginal NLS advantage observed in the original study was strengthened: we found a significant advantage for the NLS classification. These results are discussed with respect to accounts provided by formal models of human classification learning.

Linearly separable (LS) classification: members of two categories can be differentiated on the basis of a weighted, linear combination of featural information.

Prototype theories (e.g., Smith, Murray, & Minda, 1997) assume classifications follow LS structure, predict NLS classifications are more difficult.

Exemplar theories (e.g., Medin & Schaffer, 1978) are not constrained by linear separability.



Medin & Schwanenflugel (1981, E4)

Critical test of human sensitivity to linear separability.

Learners trained on LS or NLS classifications, matched on within- and between- category exemplar similarity.

Stimuli: Photographs of faces differing in hair color (dark, light), hair length (long, short), and smile type (closed, open). Different instantiations of the logical dimensions for each block (e.g. the same exact photo was never shown twice)

Training: 18 blocks unless criterion of 2 error-free blocks met

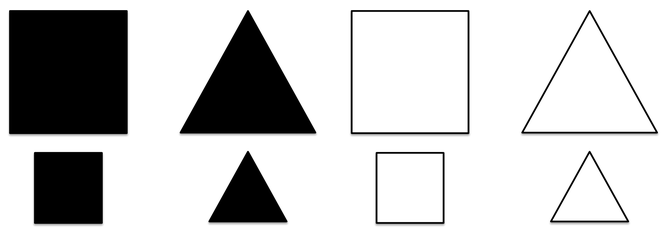
Result: Mean errors across training for LS (39.5) and NLS (38) were not significantly different.

Results cast doubt on prototype theories of classification learning. While findings were influential, conclusions are were open to the following criticisms:

- Null result perhaps obtained due to low *n* (16 per group)
- Nonstandard stimulus materials (limiting formal models)
- Low level of mastery demonstrated (only 11/32 reached criterion across the 16 blocks)

The present study is an attempt to replicate and model key findings with more standard materials, higher power, and more extensive training. Typicality ratings will provide a glimpse into category representations of learners.

Stimuli: Geometric Shapes



Assignment of logical structure to physical dimensions was fully counterbalanced.

Category Structures:

Same as Medin & Schwanenflugel (1981) Experiment 4

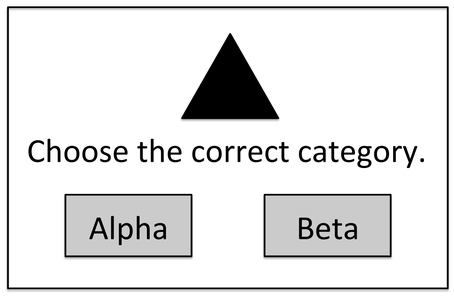
LS		NLS	
A	B	A	B
110	001	100	000
101	010	011	001
011	100	111	110

This pair of structures is an ideal test case because within- and between-category exemplar similarity is matched.

Task: Participants were trained on 25 blocks of classification learning (without a criterion), were given a test block without feedback, then were given a pairwise typicality test phase.

Training: 25 blocks

Traditional classification task with corrective feedback



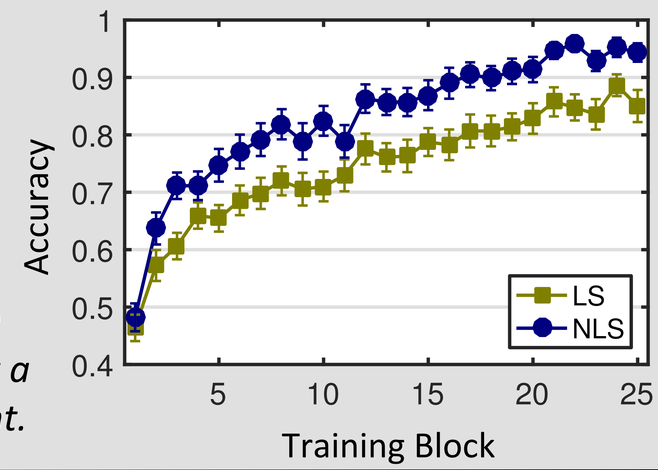
LS n = 69**, NLS** n = 56

Core result replicated

NLS classification not more difficult than LS.

NLS *less* difficult than LS

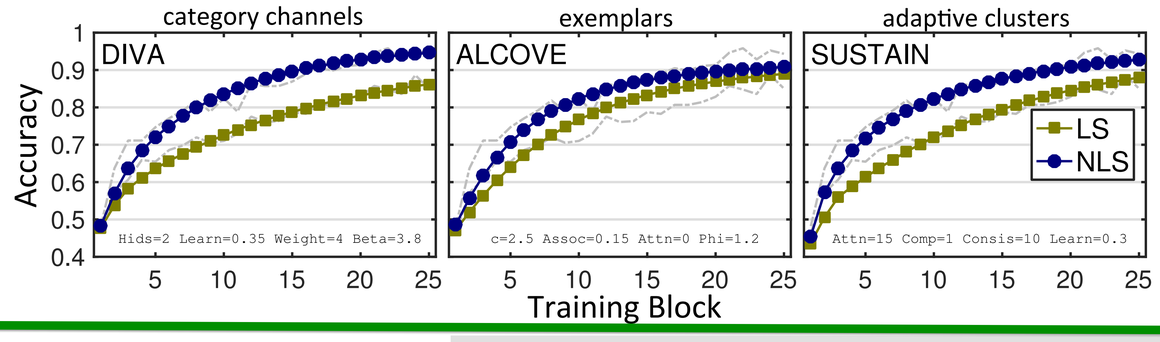
Additional evidence against a linear separability constraint.



DIVA
(Kurtz, 2008)
Classifies based on relative reconstructive success along category channels

ALCOVE
(Kruschke, 1992)
Classifies based on similarity to stored exemplars

SUSTAIN
(Love, Medin, & Gureckis 2004)
Classifies based on similarity to adaptive clusters

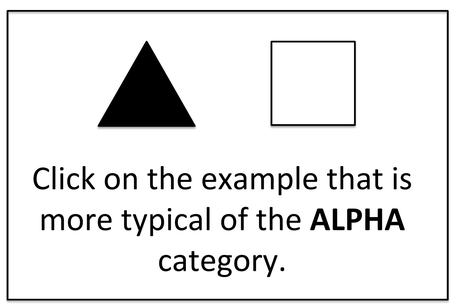


Model Simulations of NLS Advantage During Training

For each classification model, a grid search was performed to find parameters (listed) that best explain the observed performance. (*All MSEs < 0.002.*)

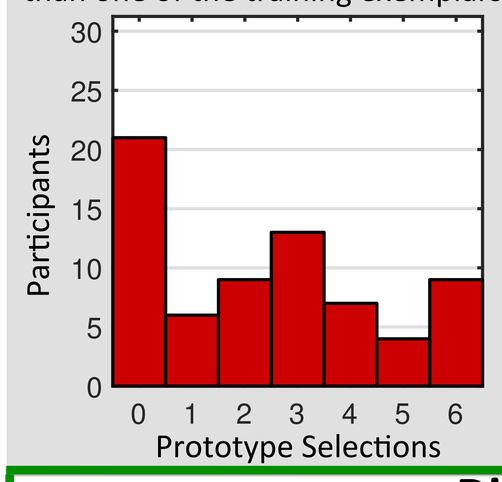
Pairwise Typicality:

Participants were asked to choose between (novel) **prototypes** and training exemplars. *i.e.,* ***000*** *vs 010*



Particularly because typicality can be interpreted as familiarity, If the prototype is thought of as more typical than training exemplars, it could indicate a **family resemblance** representation of the categories (e.g. the learner has a sense of the most typical value along each dimension).

Distribution of # of trials (out of 6) that participants chose the **prototype** as being more typical than one of the training exemplars



Individual differences in LS group

20/69 viewed the prototypes as more typical than the training items on more than half (>3) of trials.

36/69 viewed the training items as more typical than the prototypes on less than half (<3) of trials.

There was no difference in training performance between these two subgroups of LS learners, p > .05.

Discussion

- NLS structure learned faster than LS, extending findings from M & S (1981) with more subjects, more extensive training, and standard materials.
- Popular category learning models were able to account for NLS advantage.
- Typicality ratings indicated more than one type of category representation but this factor did not seem to affect learning success.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review, 85*, 207.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, 7, 355-368.

Smith, J. D., Murray, M. J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 659.