The proposed model has a couple of desirable features lacking in current models, mostly a natural but pivotal role played by learning variables known to shape concepts, e.g., variables such as category size (the number of different patterns defining a category) and pattern distortion (which fixes the breadth or boundary of a category). I believe there is a qualitative shift in processing strategy - not necessarily a deliberate or conscious act but a strategic shift nonetheless - which occurs when category size reaches a size sufficient to preclude veridical memory for an event (much like STM has a capacity that can be exceeded). I've attempted to formalize this model below - it is, as you'll see, not always a quantitative formalization.

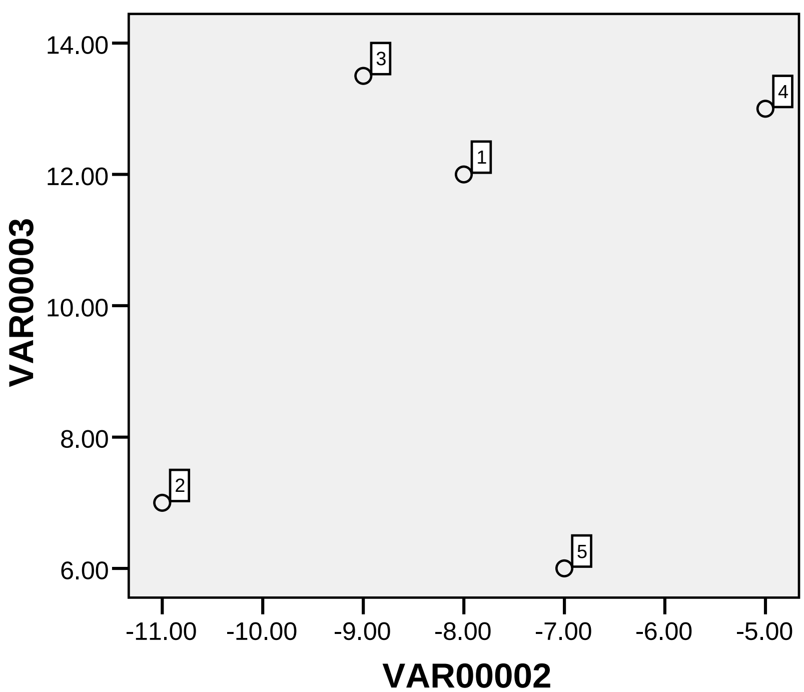
Learning

1. In learning, the subject attempts to retain in memory each pattern presented, preserved as much as possible in its presented form. With a certain probability (probably a parameter here - α), the subject attaches correctly the label (provided by learning feedback) to this pattern. Although each pattern is composed of 9 points, with points connected by lines to form connected shapes, the examples below use only a single point (with an {x,y} coordinate value) of the 9 available for simplification, at least initially.

Trial 1: Pattern 1-A_1

2. On trials 2-N, memorization proceeds as in 1 above (although illustrated only for category A, it is extendable to multiple categories), e.g.,

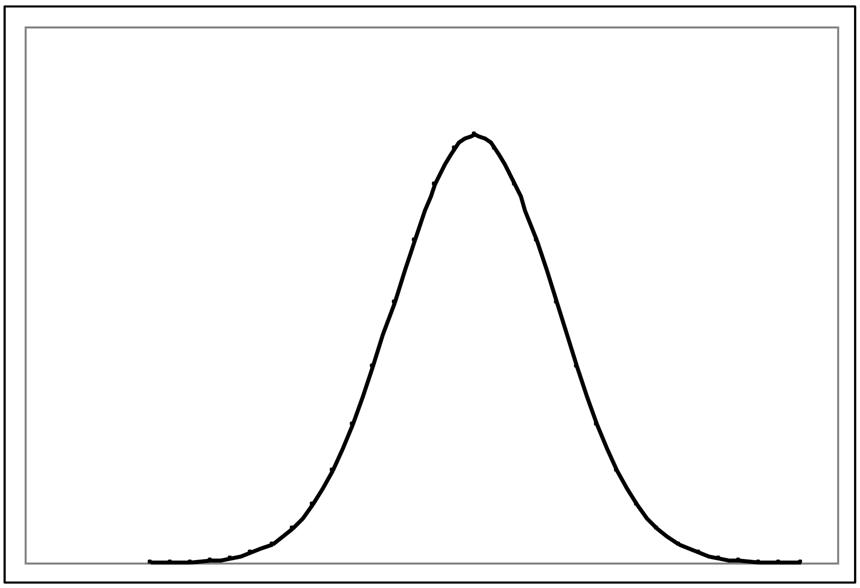
Trial 2: Pattern 2-A_2
Trial 3: Pattern 3-A_3
Trial 4: Pattern 4-A_4
Trial 5: Pattern 5-A_5



3. After a certain point, the number of patterns (N) presented for learning exceeds memorization threshold and a strategy shift occurs - rather than storing each pattern, the subject computes the mean and variance for that point, and this is stored, i.e.,

Trial N+1: $\mu_{N+1} = [N\mu_N + \{\text{pattern } N+1\}]/(N+1)$

σ_{N+1} = variance for these N+1 points



-- I assume the distribution should/could/might be a normal bivariate

4. With additional patterns defining A, the mean & variance (μ, σ) is continually adjusted.

5. Importantly, if the number of training patterns defining a category is relatively small (< N), the subject simply attempts to store each training pattern with its associated label - this is, of course, the exemplar model of classification.

(i) It is worth noting that further increases in category size, once N is exceeded, will only slightly alter μ; however, the level of training distortion can result in dramatic differences in the final state of σ. This latter value of σ can markedly affect classification on a subsequent transfer test (as shown in Homa & Vosburgh, 1976; Homa & Cultice, 1984).

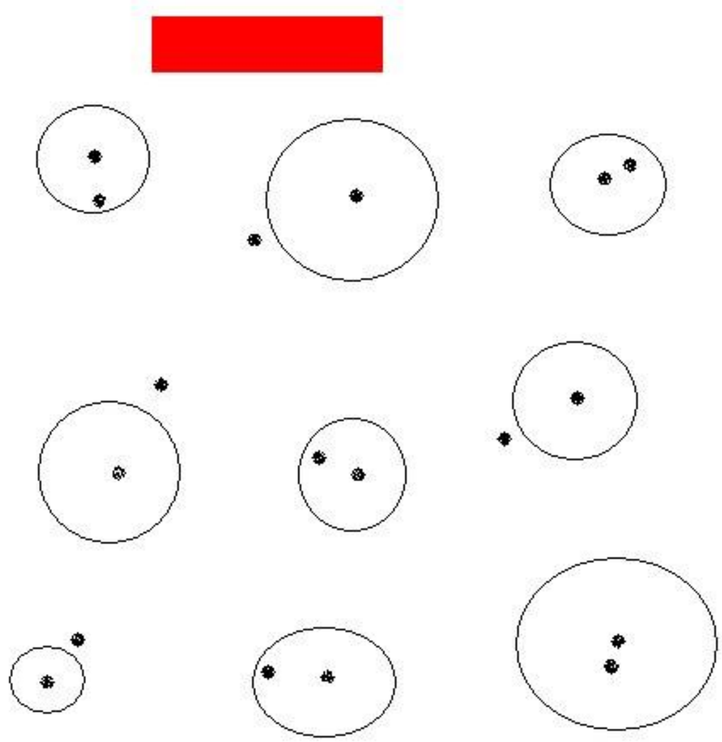
Transfer Test

On the transfer test, the subject is either asked (without feedback) to **classify** old and new patterns into appropriate categories or indicate whether a given pattern is old or new (**recognition**). These are considered separately - first, classification

Classification

1. If learning culminated with exposure of only a handful of examples, i.e., some value < N, then classification of each transfer pattern is based on the summed similarity of the presented pattern to each of the patterns stored in memory (e.g., Nosofsky's generalized context model).

2. However, if learning culminated with exposure to numerous examples, i.e., some value > N, then classification is determined by whether each point of the transfer pattern is contained within the circumscribed breadth of each stored point. Each 'point' of the stored category is represented by its mean and variance (the mean is shown as the centroid of each distributed point, and its breadth by a circumscribed region based on its variance), e.g.,



For the left-most point (circle, upper left), the point of the transfer pattern is contained within the breadth of the category - it's within σ - units of the mean; for the second point (upper row, middle circle), the transfer point falls outside the breadth. Overall, the 9-points of the transfer pattern result in 5 points within the breadth and 4 outside it.

3. Classification decision rule: A couple of possibilities, each similar to each other:

A. Find the category containing the most 'included' points, and assign the transfer pattern to that category. That is, if transfer pattern X is contained within 5 regions of A, 3 regions of B, and 2 regions of C (assuming 3 categories), then assign it to A (deterministic strategy); or

B. Compute the overall similarity (based on σ - units) of the transfer pattern to each category, based on summed similarity for each category.

C. A minor refinement would be to assign a given transfer pattern to a ‘junk’ category (or ‘none of the above’) if the overall similarity fails to exceed a criterion for any of the categories.

-- What this predicts so far:

1. Better classification when training occurs on high-variance patterns but only when the category size is large (Homa & Vosburgh, 1976 and later studies)
2. Generally supportive of our earlier work that a prototype parameter generally is manifested only following training on numerous versus few category instances (Homa et al., 1991)
3. Sharp gradients based on old-new similarity when category size is small with little effect of similarity to the category prototype; but flatter gradients based on old-new similarity when category size is large, with strong effects of prototype similarity (Homa et al., 2008)

Recognition

Recognition judgments are based on information similar to what drives classification - similarity. What is critically different, however, from existing models is that recognition is based on similarity to stored exemplars following learning, which occurs when category size (N) is small, e.g., N < 5 or so. When category size is increased, similarity is based on the only representation available - the stored prototype and variance. This outcome insures an advantage in old-new discrimination only when category size is small but not large, i.e.,

1. This would predict that old-new recognition would be good for a number of situations: when the number of training patterns is low (e.g., N < 5), especially when multiple learning trials are given and especially when variance is high (patterns should be more discriminable)

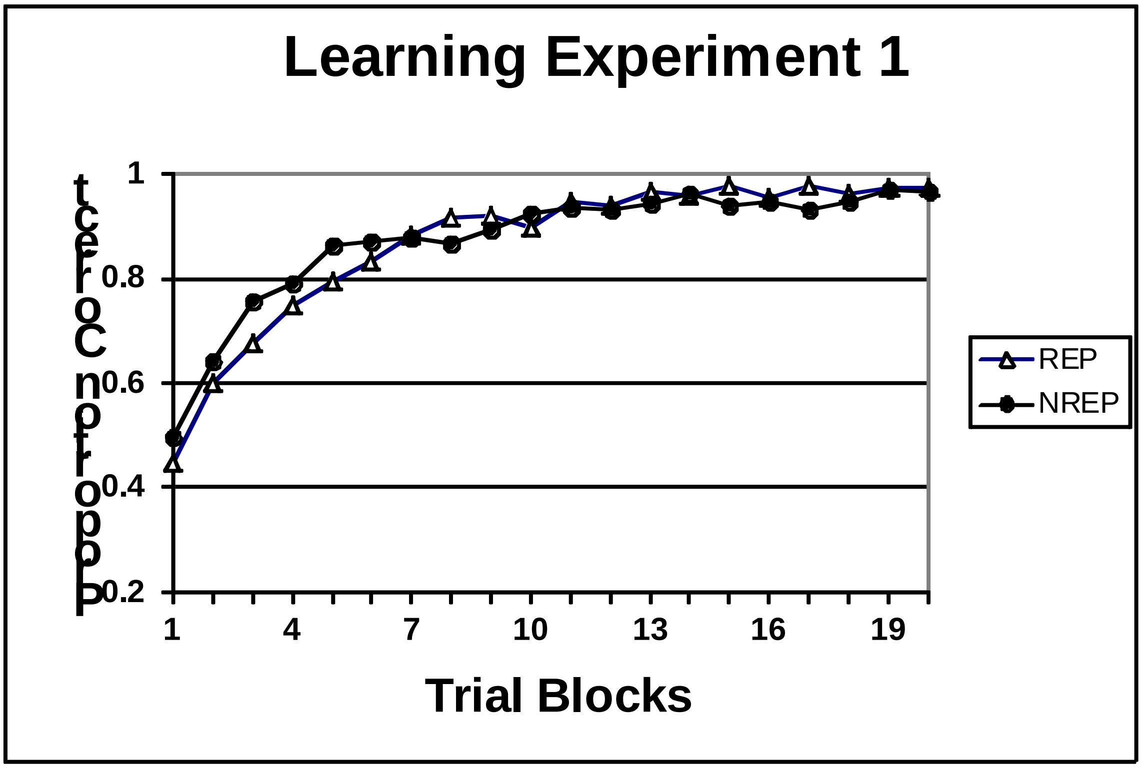
2. However, old-new recognition should deteriorate when category size is large, and the number of repetitions of training patterns in the learning phase is few or minimal.

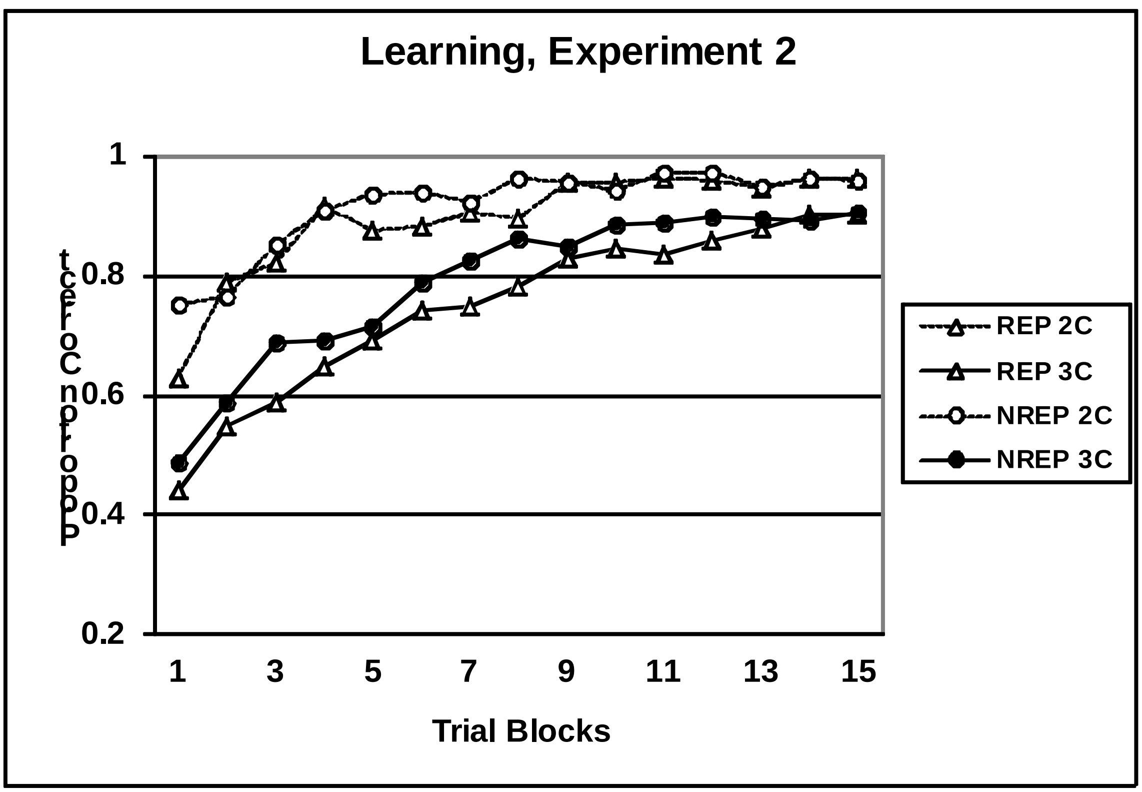
Results

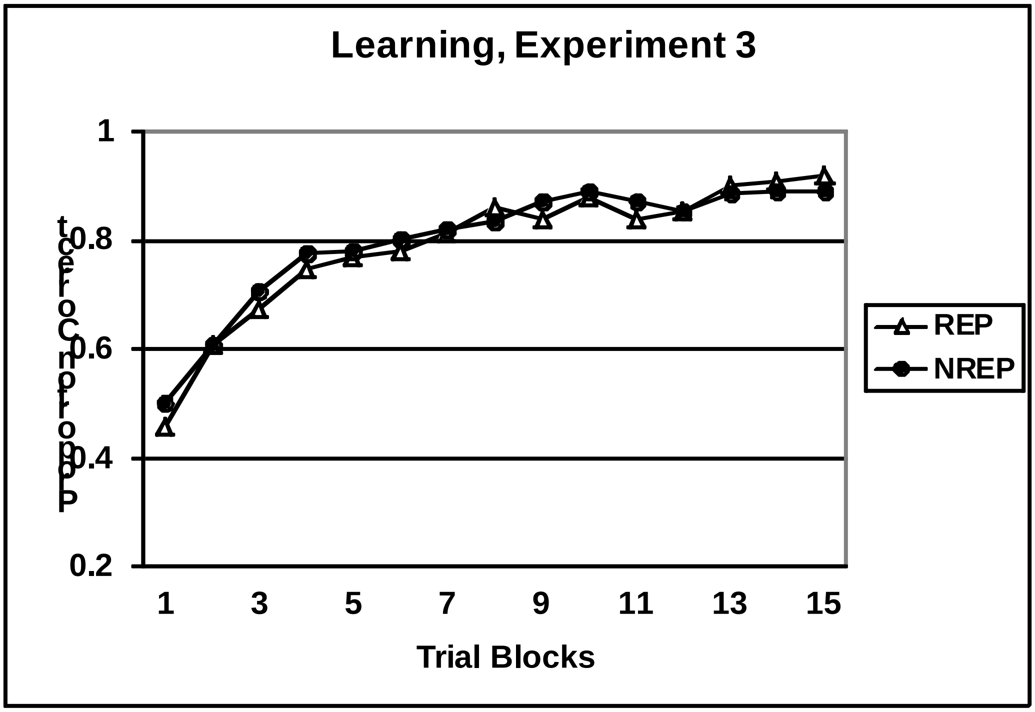
We have completed a number of studies that generally supports this model. In brief, the subject views, in learning, five different patterns that belong to either two (one condition) or three (different condition) categories. In the REP condition (repeated patterns), these patterns are repeated across 15 blocks. In the NREP condition (non-repeated patterns), the subject attempts to classify five patterns that belong to either two or three categories. Importantly, the patterns on any block are new, i.e., the subject never sees the same pattern more than once. After learning, the subject receives either a transfer test involving classification of old and new patterns or a recognition test. No feedback occurs on transfer; the new patterns include patterns at low-, medium-, or high-distortions from the prototype, as well as the prototype itself.

There are three critical pieces of data: (a) Learning across blocks; (b) Classification; and (c) Recognition. In a nutshell, we have found the following results (full details are shown in our preliminary paper – Homa & Blair – which is contained on my website (asu.edu):

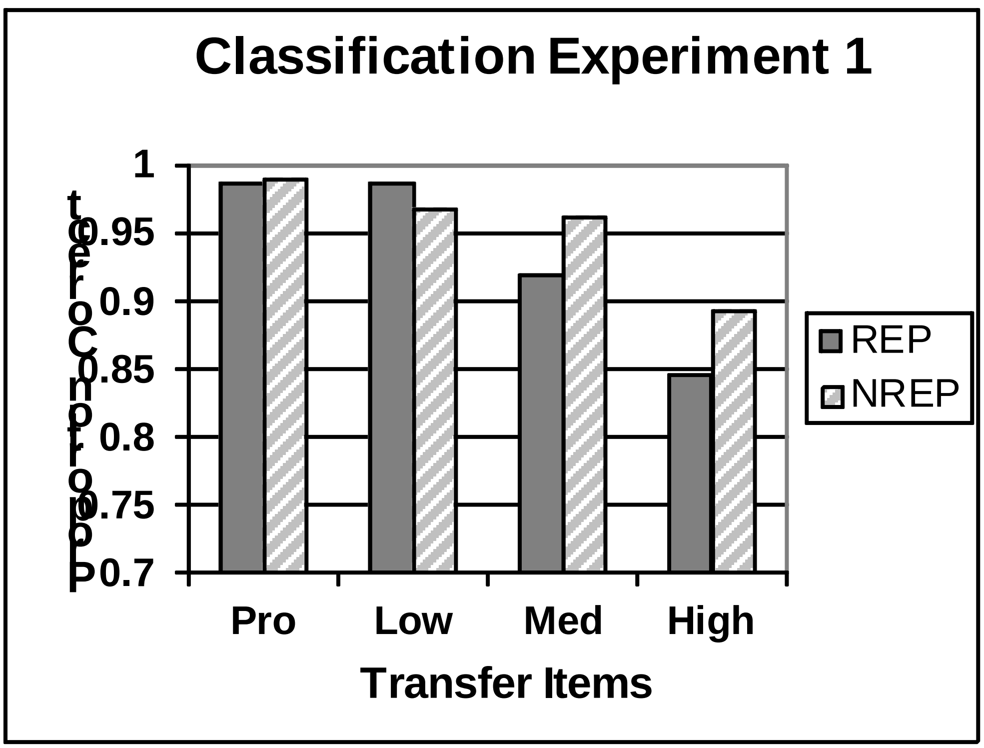
1. <u>Learning rate is unperturbed</u> by having the patterns novel on each learning block. Across three experiments, the NREP actually as a slight advantage on the early blocks; on the near terminal blocks, the REP condition catches and slightly surpasses the NREP (both groups terminate at over 90% accuracy). We suspect the REP subjects eventually memorize the five patterns in each category.


Learning Experiment 1
Proportion Correct
1
0.8
0.6
0.4
0.2
1
4
7
10
13
16
19
Trial Blocks
REP
NREP

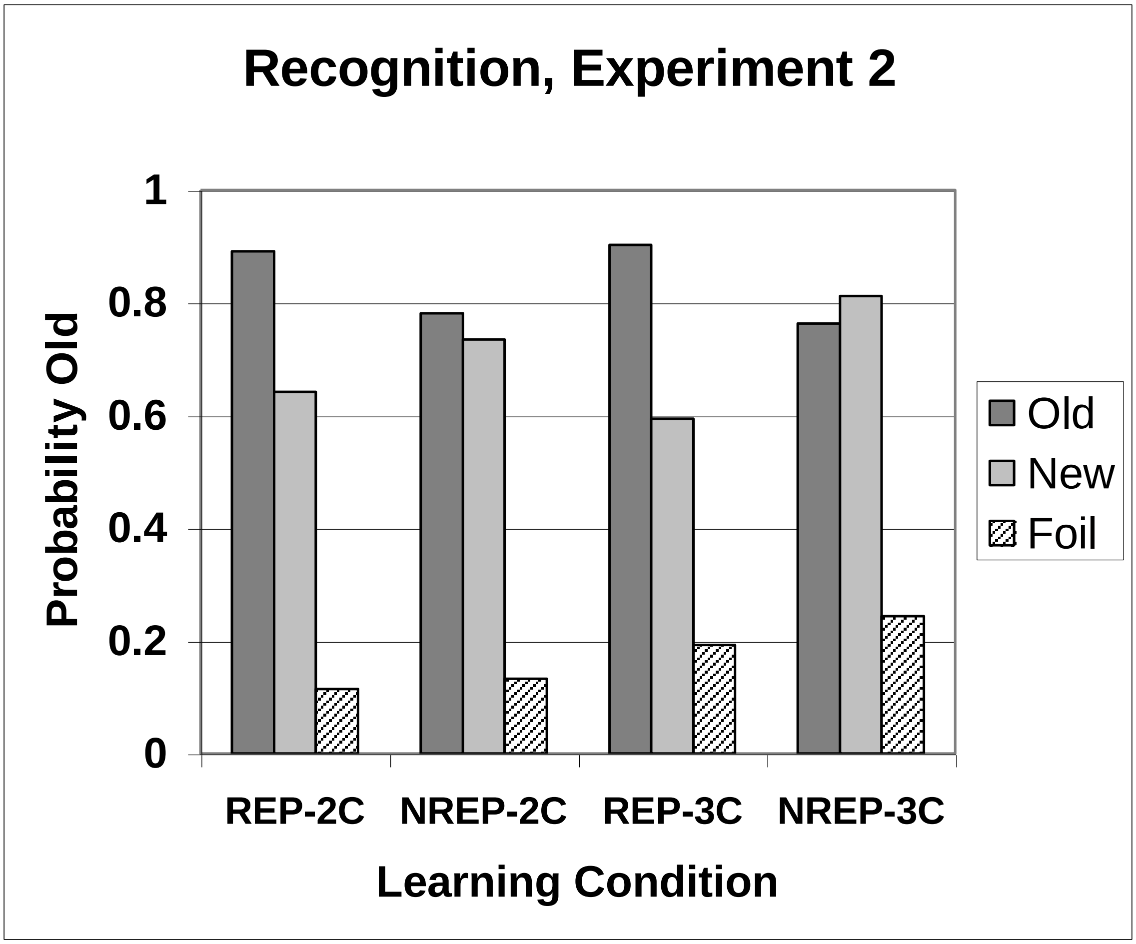

Learning, Experiment 2
Proportion Correct
1
0.8
0.6
0.4
0.2
1
3
5
7
9
11
13
15
Trial Blocks
REP 2C
REP 3C
NREP 2C
NREP 3C



2. Transfer – classification slightly favors the NREP condition, with statistically significant differences for the medium and high distortions.



3. <u>Transfer-recognition favors the REP condition</u>. In two experiments, subjects show a clear ability to discriminate old (training patterns) from new; the NREP subjects show no ability to discriminate old training patterns from new patterns at the same level of distortion.
4.



In summary, we have found the following key results which I believe this descriptive model predicts:

1. Rapid learning, with no difference in rate, between the REP and NREP conditions (actually, learning has consistently appeared to be slightly faster in the NREP until late in learning, when extended experience with patterns in the REP condition finally surpassing (slightly) performance in the NREP condition

2. Little difference in classification rate on the transfer test, with a slight advantage (significant in two cases) for the NREP condition

3. Excellent old-new recognition in the REP condition and no apparent ability to discriminate old from new in the NREP condition.

Formal Model

With assistance from John Medema (a computing specialist) and Dr. Greg Stone (a colleague skilled in formalizing quantitative modeling), we have had some success – we can successfully model the learning results, showing no disadvantage in learning for NREP (indeed, a slight advantage early on, which is what we have found) and mediocre recognition for the NREP condition. We're still in the developmental stage but this model shows promise.