# How exemplar availability affects categorization?
# Comparing categorization models.

Alberto Greco, University of Genoa

**SHORT REPORT**
**October, 1997**
**(provisional draft)**

## Abstract

An experiment was performed to test the hypothesis that each of the three classical categorization models (exemplars, features, and prototypes) applies depending on the kind of exemplars available in a specific context. Subjects rated category membership of new exemplars, on the basis of positive or negative exemplars of a category. The experiment was run in three stages, where the characteristics shared by positive exemplars, and not shared by negative ones, were manipulated in such a way that new cases would be evaluated, as expected according to each of the three theories, (1) by similarity to known exemplars, (2) by shared binary properties, (3) by shared graded properties. Results failed to conform to such pattern, as subjects exhibited logical inconsistency in evaluating cases according to available evidence. This outcome is discussed.

## Introduction

According to the three main current approaches to categorization (Smith & Medin, 1981) categories are constructed by abstraction of defining *features* (Smith, Shoben, & Rips, 1974), or by comparison of new cases with previous *examples* (Brooks, 1978), or with specially representative examples called *prototypes* (Rosch, 1975). These theories are not necessarily in contradiction, and it seems reasonable to suppose that category induction depends on the kind of available exemplars. New cases can be related to prior cases considering known *exemplars* or known *properties*.

However, to distinguish on empirical grounds between a categorization strategy based on abstraction from one based on exemplars may be difficult (Malt, 1989; Lamberts, 1994). In order to test the hypothesis that each of the three classical categorization models (exemplars, features, and prototypes) applies depending on the kind of exemplars available in a specific context, we performed an experiment, originally suggested by Wang (1993).

Subjects, on the basis of positive or negative exemplars of a category, were asked to rate on a scale how much new exemplars belonged to the category. The experiment was run in three stages:

1) given positive exemplars resembled each other and at the same time were very different from negative ones - in this case, the most perceptually similar pictures were expected to be judged as the best category exemplars; new cases would be related with known *exemplars* (extensional relationships), in accordance with the exemplar theory description;

2) all given positive exemplars had a common property (e.g. three upright rectangles) - in this case, it was expected that Ss. would use an explicitly definable rule based on this attribute, relating new cases with known *properties* (intensional relationships), as described by the classical (categorization by features) theory;

3) based on new exemplars, only a graded property (e.g. tending to be aligned) was still shared by all positive exemplars; at this stage, categorization would still be made based on an intensional relationship, not all-or-none (binary) as in the stage 2 but graded; the prototype theory here should apply better.

## Method and procedure

24 subjects participated. For task explanation purposes even/odd numbers were used as an example, and Ss. did a warm-up task with black figures as positive / white figures as negative examples, to ensure understanding. As materials for the real task, geometric figures were used (figures 1-4).

Subjects were told that P were positive and N negative examples of a category and asked to evaluate, on a 0-5 rating scale, how much each of the testing examples (X1 to X6, figure 2) was a member of the category. At stage 1 only figure 1 examples were available, at stage 2 information from figure 3 was added, and at stage 3 also figure 4 was available. During the task, Ss. had a sheet with teaching examples labeled "POSITIVE EXAMPLES" and "NEGATIVE EXAMPLES" (no other label); each subject was given one of 4 versions (each with a different order) of testing examples. There were no time limits. At stages 2 and 3 previous examples were still available and it was stressed that all of them were examples of the same category. After the whole task, we conducted an interview in order to assess the subjects' explanations of their membership ratings.

We computed the mean rating value for each of the 6 examples at the three stages. Furthermore, we coded each subject's explanation into one of the following slots:

1) similarity with positive examples (or dissimilarity with negative);

2) explicit statement of a feature (or of a combination of features) all-or-none ("binary");

3) explicit statement of graded feature or features (this code was applied if a subject used expressions like "not completely...", "they tend to be...", etc., or if a feature which logically can be graded was used - like "they are aligned" - *and* intermediate scores were used).

4) not expressable, unclear, confused.

## Results

The expectations are only partially fulfilled, and not for the expected reasons. In particular, X1 and X2 testing examples have the greatest average value (about 3.20-3.30 where 0=min and 5=max. possible values) at the first assessment stage, but X3 and X4 fail to have the greatest value at the second stage (where all figures have a value between 2.05 and 2.33), yet X4 has the *lowest* value at this stage (1.33). At the third stage, X6 has the maximum value (3.52), but X5 does not. In sum, the expected trend is not clearly confirmed.

Perhaps coding criteria were more complex than expected. 50% of Ss. started with explicit (binary) criteria, and 30% with similarity. But similarity (extensional relation) was used almost only at the first and at the third stage, however for different reasons: either when knowledge was too little or when knowledge was ambiguous or inconsistent.

Present results failed to show a direct influence of available information on categorization. As happens in many reasoning tasks, Ss. did not adopt completely rational and consistent strategies. Sometimes rules being explicitly expressed were not the ones that Ss. actually followed, which is evidence that non-conscious criteria were used in some cases.

In order to assess whether the logical inconsistency shown by subjects was due to the abstractness and artificiality of stimuli, we replied the experiment using concrete pictures of human faces (figure 5 shows testing examples), but results were very similar. (Detailed results are omitted from this version of this paper).

## Discussion

These results suggest that when generically speaking of "categorizing" it is useful to distinguish the coding process (comparing new examples with old ones, coding regularities) from the decision about category membership. When we speak of binary or graded categorization, then we are referring to the membership decision; when we speak of instances or properties, then reference is made to the coding process.

There is a possibility that results were due to an "extensional bias", that is intensional relations (induced properties) could not play a central role in the 2nd and 3rd stage as expected, since the concept is defined extensionally (by given examples), and also used extensionally (by judging membership relations). A possible solution would be to ask subjects to evaluate not example membership, but which property in a list defines better the category.

Furthermore, in order to make the task easier to be understood, instead of asking subjects to judge membership for an anonymous category, a name for it may be used (e.g. a nonsense sillable like 'DAX'). In practice, the "Positive examples" and "Negative examples" labels could be replaced with "DAX" and "NON-DAX" labels.

A revision of the experiment is planned in order to fix these points.

## References


Brooks L.R. (1978) Non-analytic concept formation and memory for instances. In Rosch E., Lloyd B.B. (eds.) *Cognition and categorization*. Hillsdale, N.J: Lawrence Erlbaum Associates, pp.169-211.

Lamberts K. (1994) Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learn.,Mem.,and Cogn.*, **20,** 5, 1003-1021.

Malt B.C. (1989) An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learn.,Mem.,and Cogn.*, **15,** 4, 539-555.

Rosch E. (1975) Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **3**, 192-233.

Smith E.E., Medin D.L. (1981) *Categories and concepts*. Cambridge, Mass: Harvard University Press.

Smith E.E., Shoben E.J., Rips L.J. (1974) Structure and process in semantic memory: a feature model for semantic decisions. *Psychological Review*, **81**, 214-241.

Wang P. (1993) Comparing categorization models. A psychological experiment. *Technical Report* n. 79, *Center for Research on Concepts and Cognition*. Indiana Univ., Bloomington.

Figure 1: Teaching Examples I

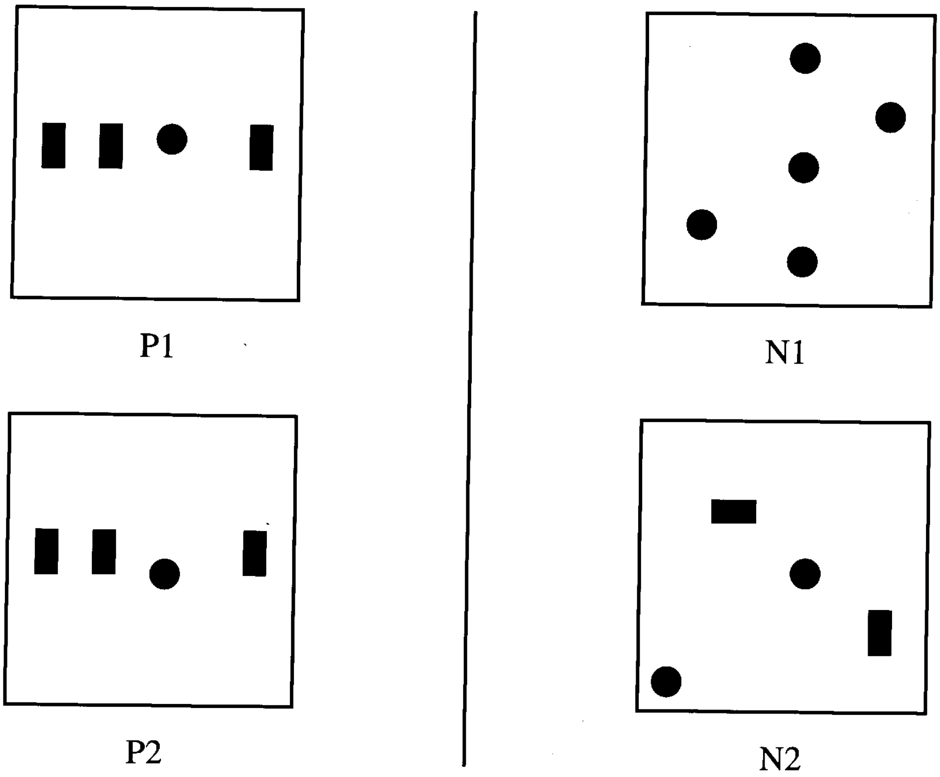

Figure 2: Testing Examples

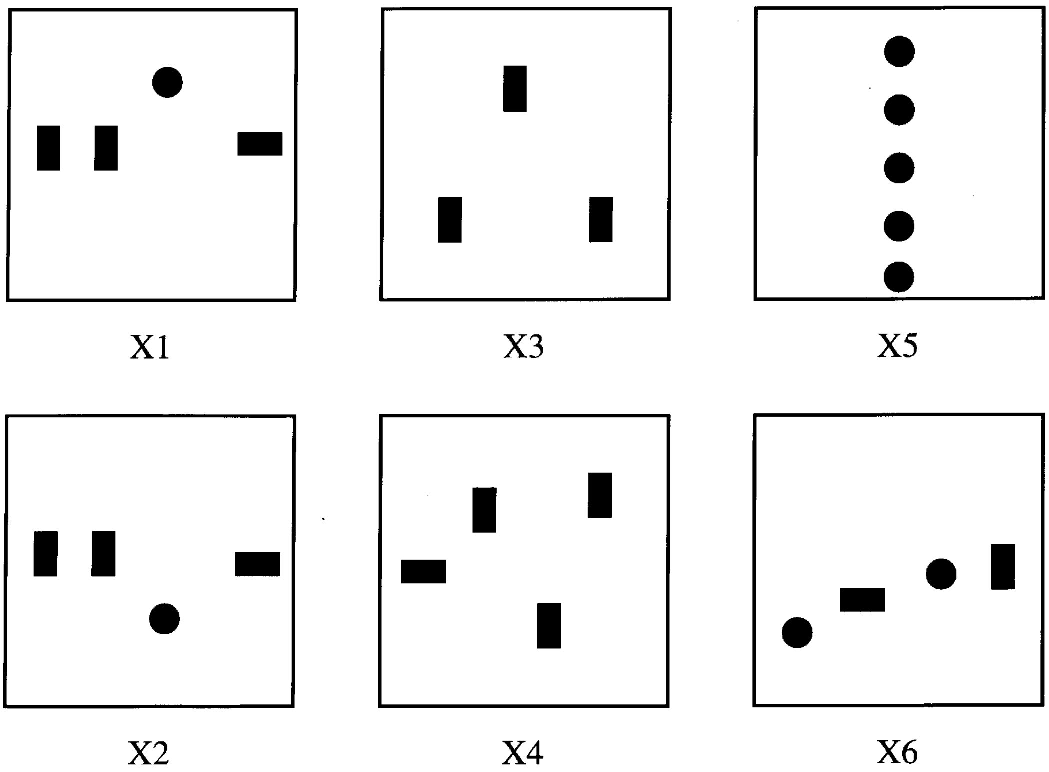


Figure 3: Teaching Examples II

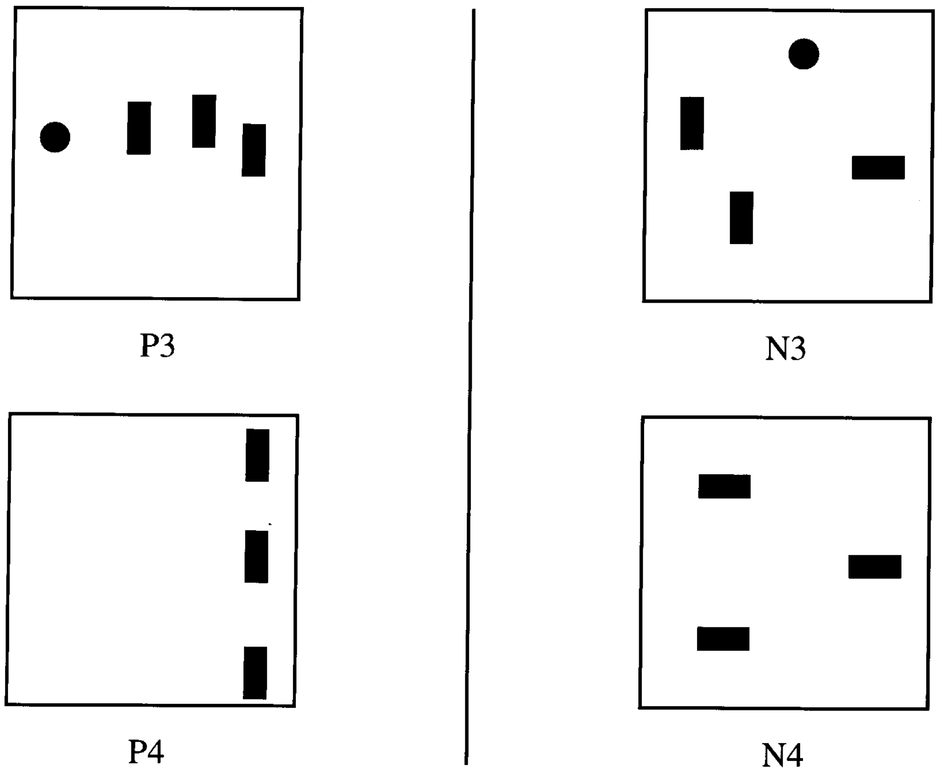

Figure 4: Teaching Examples III

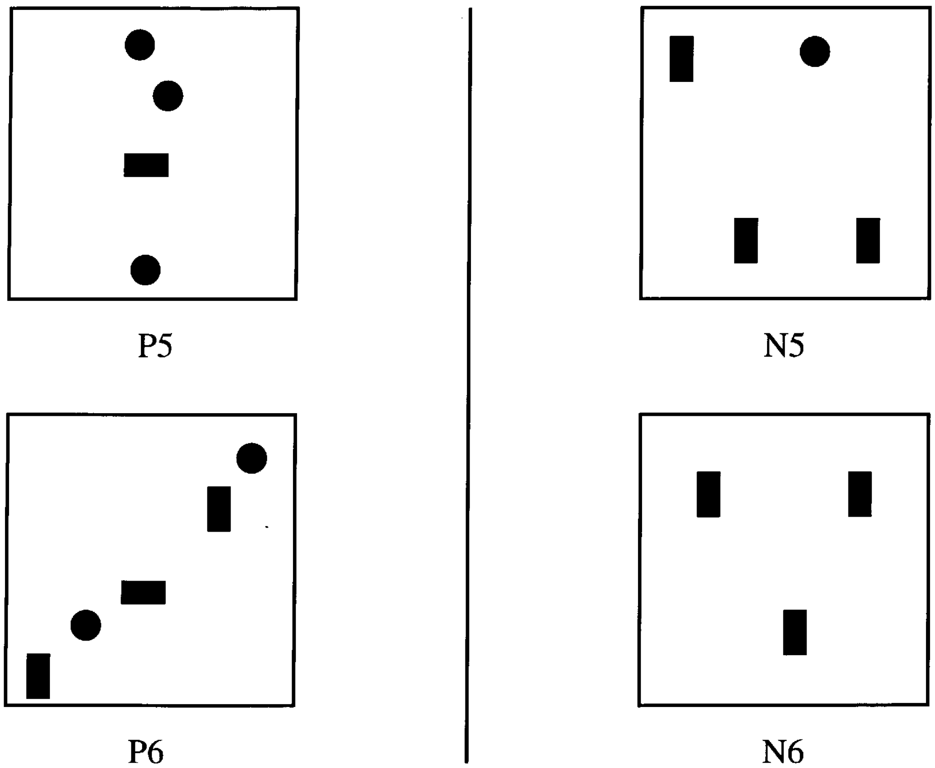


Figure 5: Experiment II - Testing Examples