UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Building The New Onto The Old: Category Constraints on Category Formation

Permalink
https://escholarship.org/uc/item/7r40x67r

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 20(0)

Author
Cabrera, Angel

Publication Date
1998

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Building The New Onto The Old: Category Constraints on Category Formation

Ángel Cabrera (acabrera@inf.uc3m.es)
Departamento de Informática
Universidad Carlos III de Madrid
Butarque 15
28911 Leganés (Madrid), SPAIN

Abstract

It is generally accepted that the process of forming a new category is biased by the learner's prior knowledge. In this context, numerous studies and models have paid attention to the effects of prior domain theories on the process of forming new categories. What is yet to be understood is how this process of acquiring new knowledge might be affected by background knowledge of the very same type, i.e. by prior categories. This paper presents a few experiments showing how the formation of new categories might be facilitated by a high *overlap* between the new and the old categories, where *overlap* is operationalized as the mutual entropy between the two.

Introduction

The idea that the acquisition of new knowledge is biased by the learner's prior knowledge is one of the basic tenets of cognitive science with respect to learning phenomena. The idea goes back, at the very least, to the Piagetian notion of *assimilation,* according to which new knowledge is constructed onto, and consequently constrained by, existing knowledge. Piaget stressed the idea that background knowledge acts as a lens through which meaning is attributed to one's observations, in a way that helps the cognitive system maintain a high degree of internal coherence or equilibrium.

Since Piaget, several researchers have explored the biasing (or guiding) role of background knowledge in learning. Just to mention a few, Wisniewski and Medin (1994), and Pazzani, (1991) have studied the interactions between prior theories and observation in the formation of new categories. Murphy and Medin (1985) attribute prior theories the important role of "holding cateogries together", a role that can also be traced to Barsalou's work on ad-hoc categories (1983) and to studies of children's conceptual development (Keil, 1989). In addition to prior theories, other forms of background knowledge might also have similar effects. Cabrera and Billman (1996), for example, found that knowledge of one's language can bias the process of category formation in a way that is not very different from the effects of prior theories.

On the machine learning side of cognitive science, several models have been proposed that deal with the effects of theory-driven inductive process in learning (e.g. Mooney, 1993; Mitchell, Keller & Kedar-Cabelli, 1986; DeJong & Mooney, 1986).

In all these cases, the assumption is that one type of prior knowledge (domain theories or language) affects the formation of knowledge of a different kind (concepts or categories). In contrast, this paper focuses on how prior knowledge of one kind (categories) can bias the formation of knowledge of that very same kind.

Categories as mediators of generalization

One way to start analyzing how existing categories might influence the formation of new ones consists of looking at the role played by categories in generalization. Categories have been shown to play an important role in induction by acting as conducting surfaces spreading the generalization of new knowledge from one object to a whole class of objects. In an illustrative developmental study by Gelman and Markman (1987) children showed a strong tendency to generalize new properties they had learned about one object to all the members of that object's basic-level category. Furthermore, when the categories were made more conspicuous by being explicitly labeled, the effect was reinforced, and category mediated generalizations outnumbered generalizations driven by the appearance of the objects. Other examples of category mediated generalizations can be found in the area of language development. For instance, in the acquisition of the past tense, children find no trouble in generalizing morphological rules such as those required to form the past tense to all members of the VERB category (e.g. "she jump<u>ed</u> very high", "he go<u>ed</u> to bed") and to no other classes of words (see Pinker, 1989, for a discussion).

So, if categories constitute a milieu for the generalization of new knowledge, it would seem reasonable to expect that they also help guide the generalization of membership to a new category being formed, since category membership is not but a particular kind of knowledge. In other words, if I learn that Lassie is a MAMMAL, I can use my prior category DOG to generalize this new piece of (categorical) knowledge to all dogs. If this interpretation is correct, we might be able to look at existing categories as conducting surfaces spreading the generalization of new category membership, a view that will help us understand the process of building new categories over old ones.

Overlap and Mutual Entropy

According to this view, the formation of new categories will be facilitated if the new categories tend to respect the boundaries of existing categories, i.e. if they tend to hold together items that also belong together according to the existing categories. Conversely, if the new categories require distinctions not present in prior categories, their learning might be hindered.

This idea of "agreement" or "disagreement" between alternative categorizations of a domain can be captured by a construct that I will refer to as *overlap*. From all the possible operational definitions of *overlap* I have chosen one that borrows from information theory, a framework with a successful predictive record in the area of human categorization (Corter & Gluck, 1992; Cabrera, 1995). The definition goes as follows.

Let us refer to each possible way of partitioning a domain as a Π. A partition Π consists of a set of exhaustive, mutually exclusive classes of objects, $\Pi = \{C_j\}$. Let $\Pi_1 = \{C_1, C_2, C_3, ..., C_n\}$, be a partition of the domain, and let $P(C_i)$ be the (prior) probability of occurrence of each class within the domain. Information theory defines the entropy associated with the set of $P(C_i)$'s as $-\sum_{i=1}^{n} P(C_i)\log_2 P(C_i)$, an expression that represents the average amount of uncertainty (measured in bits), associated with deciding which class any given instance belongs to.

For example, a partition consisting of two equi-probable classes has an entropy of 1 bit, whereas a partition of 8 equi-probable classes will convey an entropy of 3 bits. Intuitively speaking, entropy represents the average number of binary questions one would need to ask in order to figure out which category any given object belongs to. In general, the more classes in a partition, the more uncertainty the partition will convey. Also, the closer the probabilities are to being uniformly distributed the higher the uncertainty.

Now suppose that we have two partitions, Π_1 and Π_2, where Π_1 consists of a set of n classes B_j and Π_2 consists of a set of m classes C_i. Saying that Π_1 and Π_2 have a high overlap amounts to saying that knowing the class membership of an object according to one of them reduces significantly our uncertainty (entropy) about the class membership of the object with respect to the other one. Overlap can therefore be measured as the average reduction in uncertainty that one partition provides with respect to the other:

$$\begin{aligned} Overlap(\Pi_1, \Pi_2) = -\sum_i P(C_i)\log_2 P(C_i) + \\ + \sum_j P(B_j) \sum_i P(C_i|B_j)\log_2 P(C_i|B_j) \end{aligned} \tag{1}$$

where $P(C_i|B_j)$ represents the probability of an instance belonging to C_i given that we already know that it belongs to B_j. This expression is known by information theorists as *mutual entropy*. The first term in this equation represents the *prior entropy* associated with partition Π_1, or Π_1's intrinsic uncertainty in the absence of any additional information. The term inside the brackets, which we can call *conditional entropy*, represents the uncertainty that is left with respect to partition Π_1 conditionally upon knowing the class membership of the object with respect to Π_2. This conditional entropy is weighed by the probability of occurrence of each of the categories B_j, $P(B_j)$, and averaged across all the categories in Π_2. Altogether, *overlap* represents how much more certain I am about the class of an unknown stimulus according to one partition when I know the class the stimulus belongs to according to the other partition.

The "overlap" operator can be thought of as a dot-product between partitions, representing the size of the "projection" of one over the other. Mutually uninformative partitions have an overlap of zero bits, a situation that, following dot-product terminology, can be referred to as *orthogonality*. Saying that partitions Π_1 and Π_2 are orthogonal amounts to saying that $P(C_i|B_j) = P(C_i)$ and $P(B_j|C_i) = P(B_j)$ for any $C_i \in \Pi_1$ and $B_j \in \Pi_2$.

A situation of particular interest when dealing with human categories is what we can call *embedded partitions*. We say that partition Π_1 is embedded in Π_2 (or that Π_2 embeds Π_1), and we denote it as $\Pi_1 \subseteq \Pi_2$, if for every category $C_i \in \Pi_1$, there is a category $B_j \in \Pi_2$ such that $C_i \subseteq B_j$. Embedded partitions are at the opposite extreme from orthogonal partitions in terms of overlap. Between these two extremes there is a continuous range of possible degrees of overlap between partitions.

Notice that claiming a learning bias towards highly overlapping sets of categories implies a preference for hierarchically organized systems of categories over systems of categories which cut across each other's boundaries (orthogonal organization). This prediction is consistent with the often noted hierarchic character of human categories (e.g. Mervis & Rosch, 1981; Keil, 1983).

Experimental Evidence

This section summarizes the results of three experiments that were conducted in order to test this hypothesized human preference for highly overlapping sets of categories. In each of the experiments, subjects first learned to classify a set of 16 Phoenician characters[1] into a number of categories following the feedback signals provided by the experimenter. Immediately after, subjects had to learn to classify the same items in a different way. The experiments varied the amount of overlap between the first and second sets of categories and measured the experienced difficulty in the second task.

Experiment 1

In the first experiment, 38 Georgia Tech undergraduate students underwent two successive supervised classification tasks of 240 trials each: one task with two categories and the other one with four. At each trial, subjects were presented for 500 ms with a Phoenician character inside an

[1] The stimuli were intended to be as novel to the subjects as possible in order to control for possible contaminating effects of prior theories. Similarly, categories were formed by grouping characters as independently as possible from their appearance in order to control for possible contaminating effects of topological structure.

imaginary 60 by 60 pixel square in the center of a Macintosh high resolution 13' color monitor and had to classify the character by pressing a given key in the computer keyboard. If they did not know how to classify the character they could just guess and proceed with the next trial.

In the two-category tasks (labeled as "A" and "B" in Figure 1), subjects were told that some Phoenician characters come from Persian and others come from Egyptian, and that they were required to figure out, based on the provided feedback, what characters come from what language, by pressing either "E" (for Egyptian) or "P" (for Persian). After their response, the incorrect label was removed from the screen, and the character was displayed again flanked by the correct label and the message "Correct !" or "Wrong." Subjects were instructed to respond as fast as they could while trying to be maximally correct. After responding, they could wait as long as they wished before pressing the space bar to move on to the next trial.

In the four category tasks, the trials were identical, except that subjects were told that Phoenician writers used to store their typesettings in four different boxes and that their goal was to figure out what letters went into what boxes by pressing either "4", "5", "7" or "8" on the numeric keypad.

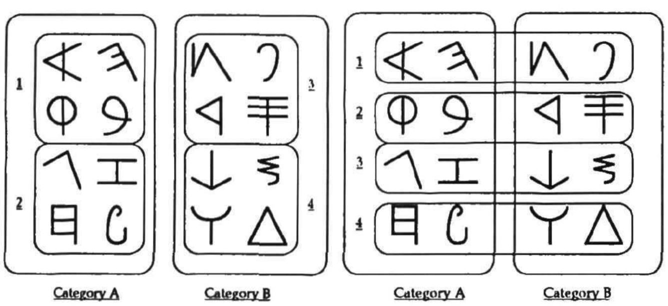

Figure 1. Categories in Experiment 1.

Subjects were split into two experimental groups. In one group (*2-4*), the four categories in the second task (see Figure 1, left panel) were embedded in (hierarchically related to) the two categories in the first task (Overlap = 1 bit according to Eq. 1). In the other group (*2-4x*), the four categories in the second task (Figure 1, right panel) were orthogonal to the two categories in the first task (Overlap = 0 bits). According to the hypothesis, experience with the first task should make the second task in the embedded situation easier than in the orthogonal situation.

Results. Two subjects in the *2-4* condition performed two SD's below the mean in the first task and were discarded for further analyses, thus leaving 18 subjects in each of the two conditions. The analysis of variance of the learning curves resulting from grouping trials in blocks of 20 yielded significant main effects of block, $F(11, 374) = 192.39$, $p = .000$, and task (first and transfer in both conditions), $F(1, 34) = 43.10$, $p = .000$. The main effect of block reflects the learning progress across trials in all conditions. The effect of task reflects the different baseline (chance) performance between the two- and the four-category tasks (chance level was 50% in the two-category tasks and 25% in the four-category tasks). There was also a significant task by block interaction, $F(11, 374) = 10.65$, $p = .000$, which reflects the fact that the effects of task vanished with block, as subjects approached perfect performance in all conditions.

According to the hypothesis, an interaction was expected between task and condition, that would show a facilitation from hierarchy in the transfer, four-category task. However, this crossover was too small to reach statistical reliability, $F(1, 34) = 1.02$, $p = .32$. The analysis of variance also showed no main effect of experimental condition, $F(1, 34) = .00$, and no significant interaction of block with either experimental condition or task.

In order to test the possible effects of individual learning ability, subjects in the *2-4* and *2-4x* were split into two groups according to the total number of correct responses during the first task. Statistical analyses of this interaction were conducted by including a dichotomized variable representing learning ability with respect to the group median (9 fast learners and 9 slow learners in each condition). The results showed a significant interaction between condition and learning ability, $F(1, 32) = 5.72$, $p = .023$, showing that fast learners and not slow learners showed the predicted benefit from hierarchical structure.

Distribution of Errors. Finally, subjects' use of their prior categories in generalizing the knowledge acquired during the transfer task was assessed by looking at the distribution of erroneous responses produced by subjects to stimuli from each of the categories in the transfer task of the hierarchical *2-4* condition. This distribution was compared to the responses of subjects from a different task who faced the same categories with no prior experience. This was done for slow and fast learners separately, in order to simultaneously assess whether the use of the prior categories in generalizing new knowledge was affected by degree of learning of the prior categories. Figure 2 summarizes the results, expressed as the proportion of incorrect responses that were consistent with the superordinate category membership of the stimulus, for each of the four categories.

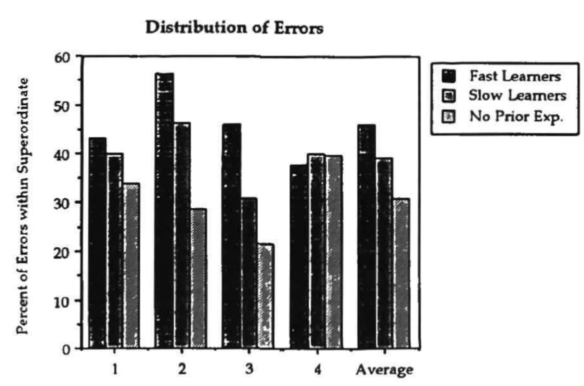

Figure 2. Distribution of errors in Experiment 1

In an unbiased situation, only 33% of the incorrect responses should correspond to the responses associated with members of the same superordinate category, because for every stimulus there were always three possible "incorrect" responses. Subjects with no prior experience performed very close to this chance level (average of 31%). Slow learners with prior experience with the superordinate categories generalized across the superodinates with a slight higher probability (39%), and more so fast learners, who averaged 46%. So, not only did subjects seem to be using the categories formed during the first task to guide the formation of categories during the second task, but also the capacity of those categories to mediate the generalization of new knowledge was more noticeable the better had the first categories been learned.

Experiment 2

The goal of this experiment was to further test whether degree of overlap with prior categories affects learning difficulty in a setting that is neither strictly hierarchical nor strictly orthogonal, but somewhere in between. In addition, this experiment introduced a new experimental design that allowed for more powerful within subjects analyses.

The categories used in each of the tasks are depicted in Figure 3. In the *AB* task there were two categories containing half of the stimuli each, i. e. $P(A) = P(B) = .5$. In the *LO* task[2], one category, *L*, contained all eight characters in category *A* plus half of the characters in *B*, while the other category, *O*, contained the remaining half of *B*. In other words, $P(L) = .75$, and $P(O) = .25$. The *AB* category set had an uncertainty of 1 bit, the *LO* set had an uncertainty of .81 bits, and the two sets overlapped in .31 bits. The *AB* task was based on the Persian-Egyptian story line of the previous experiment. The story line in the *LO* task was about how Phoenician writers organized their type-settings, keeping some in one box and others in a different box.

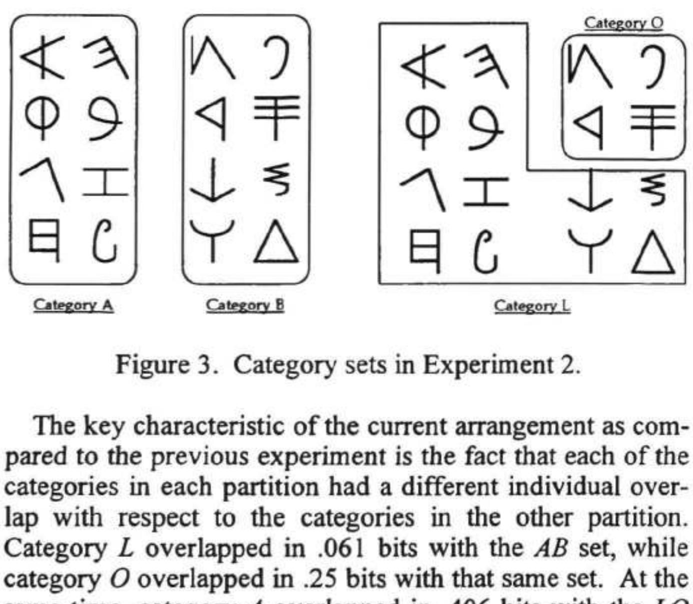

Figure 3. Category sets in Experiment 2.

The key characteristic of the current arrangement as compared to the previous experiment is the fact that each of the categories in each partition had a different individual overlap with respect to the categories in the other partition. Category *L* overlapped in .061 bits with the *AB* set, while category *O* overlapped in .25 bits with that same set. At the same time, category *A* overlapped in .406 bits with the *LO* set, while category *B* had a **negative** overlap of -.094 with that set. Interestingly (and not so intuitively!) this negative overlap reflects the fact that we are less certain about whether an instance belongs to *L* or *O* when we know that it belongs to B than when we know nothing at all about the instance. If we do not know anything about the instance, we can only guess that it belongs to L more likely than to O (probability of .75 vs. .25), whereas, if we know that the instance belongs to B and we use that information to guess whether it belongs to L or O, we are in a situation of maximum uncertainty (.5 vs. .5).

There were two experimental conditions. In condition *AB-LO* subjects experienced the AB task prior to the *LO* task. The opposite order took place in the *LO-AB* condition. According to the hypothesis, it was expected that, after forming categories *L* and *O*, learning *A* would be easier than learning *B*. Analogously, when transferring to *LO* from the *AB* task, the learning of *O* was expected to be facilitated relative to *L*. However, the effect in the latter case could be less noticeable than in the former case, given the relative values of overlap.[3] In both cases, comparisons between categories in the same task were carried out as within subjects.

There were 192 *AB* learning trials and 192 *LO* trials. The trials were randomly ordered in blocks of 48 trials. Within each block, each of the 16 characters appeared three times. Twenty-four subjects were randomly assigned to condition *AB-LO* and twenty-two to condition *LO-AB*.

Results. Across all conditions and categories, subjects averaged between 73% and 86% correct responses in the first learning task. Subjects who scored at least two standard deviations below the mean for either of the two categories in the first phase were excluded from further analyses. This was the case with three subjects in the *AB-LO* group and two in the *LO-AB* group.

<u>*LO-AB* Transfer</u>. The analysis of variance of accuracy yielded a significant main effect of condition, $F(1, 39) = 5.12$, $p = .029$, and block, $F(3, 117) = 130.84$, $p = .000$. However, the predicted condition by category interaction was only marginally significant, $F(1, 39) = 2.92$, $p = .096$.

The predicted interaction was more obvious in terms of response time. As expected, experience with the *LO* task made responses to instances from the *A* category faster than responses to instances from the *B* category, even though category *A* produced slower responses in the absence of prior experience. The analysis of variance yielded a significant main effect of condition $F(1, 39)= 8.10$, $p = .007$, and block, $F(3, 117)= 39.00$, $p= .000$, a significant condition by block interaction, $F(3, 117)= 7.72$, $p= .000$, and, most importantly, a significant condition by category interaction, $F(1, 39)= 9.63$, $p= .004$ that is consistent with the predictions.

<u>*AB-LO* Transfer</u>. It was expected that prior experience with the *AB* would facilitate the learning of *O* relative to *L*, although the effect was expected to be smaller than in the previous scenario. Statistical comparisons of accuracy in the *LO* task with and without prior *AB* experience yielded a significant main effect of experience, $F(1, 39) = 5.96$, $p = .019$, block, $F(3, 117)= 118.64$, $p= .000$, and category, $F(1, 39) = 71.53$, $p = .000$, but no significant interaction, F(1,

[2] The labels *L* and *O* represent, roughly, the shapes of the Venn diagrams of the respective categories they refer to.

[3] Notice that *Overlap(A, LO)* - *Overlap(B, LO)* > *Overlap(O, AB)* - *Overlap(L, AB)*.

39) =.18. Analogous analyses performed on response time data did not show the expected condition by category interaction either, $F(1, 39)=.79$.

Experiment 3

Experiment 3 was very similar to Experiment 2 but was based on a different arrangement of categories. The categories used in this experiment did not differ in overlap as much as those in the previous experiment but they had the advantage of allowing us to examine the distribution of subjects' errors (as we did in Experiment 1) by including a task with three categories instead of two (which gave subjects two possibilities for error in each trial).

The arrangement of categories that was used in this experiment is shown in Figure 4. Again there was a symmetrical two category task that I will refer to as *AB*. The other task, *123*, involved three categories of comparable sizes ($P(1) = P(3) = 5/16$; $P(2) = 6/16$), that were not embedded in the *AB* set but overlapped highly with it. The uncertainty associated with the *AB* set was 1 bit, the uncertainty associated with the *123* set was 1.58 bits, and the overlap between the two sets was .625 bits.

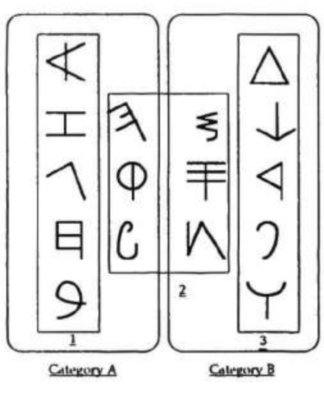

Figure 4. Design of categories in Experiment 3.

Categories A and B had an overlap of .3125 bits each with respect to the 123 set. Categories 1 and 3 overlapped also in .3125 bits with respect to the AB set, while category 2 was orthogonal to this set (zero overlap). Consequently, it was expected that experience with the *AB* task would facilitate the learning of *1* and *3* as compared to *2*. Conversely, prior experience with the *123* task should not alter the relative difficulty of *A* and *B*. These hypotheses were tested by comparing relative performance on each category between the groups of subjects with and without prior experience with the alternative category set.

There were 45 subjects. Twenty-four of them were randomly assigned to the *AB-123* condition, and twenty-one were assigned to the *123-AB* condition. The rest of the design was identical to experiment 2.

Results. Across all conditions and categories, subjects averaged between 75% and 84% correct responses in the first learning task they experienced. Three subjects in the *AB-123* group and three subjects in the *123-AB* group scored less than two standard deviations below the mean for at least one of the categories. The data from these subjects were excluded from further analyses. This left 21 and 18 subjects in each condition.

AB-123 Transfer. According to the between-within ANOVA, the expected interaction fell short of reaching the normative 5% reliability level, $F(2, 74) = 2.52$, $p = .087$. The main effect of experience also approached statistical significance, $F(1, 37) = 3.04$, $p = .09$, thus indicating that there might have been some practice effect. Finally, there was a significant effect of category, $F(2, 74) = 8.40$, $p = .001$, with category *3* leading to the overall highest scores (85.4%, SD = 5.3), and category *1*, to the lowest (80.6%, SD = 8.3). Neither the main effect nor the interaction were significant in terms of response times.

Distribution of Errors. Figure 5 (left) shows the probability of erroneously assigning to category *1* an instance of category *2* that also belonged to *A*, for subjects with and without prior experience with the *AB* task. Subjects with no prior experience with the *AB* set erred equally frequently by classifying instances of *2* as *1* and *3* (averaging about 50% to each category). On the contrary, subjects with prior experience with the *AB* task tended to err more frequently by assigning to *1* the characters that also belonged to *A*, consistent with the membership of the instances during the first task. Overall, category-consistent errors were 60% following *AB* training, but just 50%, or chance, with no prior *AB* training.

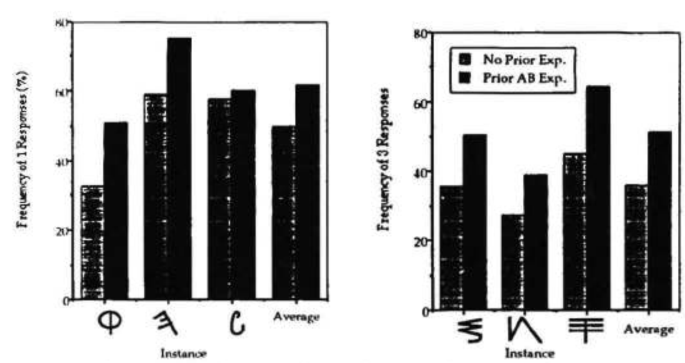

Figure 5. Distribution of errors in Experiment 3.

Similarly, Figure 5 (right) shows the probability of erroneously assigning an instance of category *2* that also belonged to *B*, to category *3*. For all three instances, the probability of being erroneously assigned to *3* was higher for subjects with prior experience with the *AB* task. These results are consistent with the predictions and the findings of Experiment 1.

123-AB Transfer. Because *A* and *B* have identical overlap with the 123 category set, prior experience with the 123 task was not expected to affect the relative difficulty of learning of *A* and *B*. The results did not contradict this prediction. A mixed ANOVA showed no significant experience by category interaction in terms of either accuracy, $F(1, 37) = .00$, or response time, $F(1, 37) = .44$. Accuracy did show a significant main effect of experience, probably indicating practice effects. There was no main effect of category in terms of accuracy or response time.

Discussion

The results of Experiment 1 showed that fast learners had less trouble transferring from a two-way classification task to a four-way task when the categories in the four-way task were embedded in (were hierarchically related to) the categories in the two-way task, than when the category sets were orthogonal to each other. On the contrary, slow learners did not show such a pattern. When looking at the distribution of errors evidence was found supporting the idea that all learners (but even more so fast learners) had a tendency to generalize category membership during the second task across the categories formed during the first task.

Experiment 2 showed that people transferring from an asymmetric two-way classification task to a symmetric two-way task had less trouble forming the category with the highest overlap with respect to the initial two categories, thus supporting the predictions. This result indicates that overlap may affect category formation even in intermediate, neither fully hierarchical nor fully orthogonal situations. When the sequencing of the two tasks was reversed, the predictions were more modest as far as possible effects of overlap, but the results failed to yield significant effects.

When subjects in Experiment 3 transferred from a three category task to a symmetric two category task, overlap predicted no effects. Not willing to support the null hypothesis based simply on a lack of statistical significance, it is of interest to note that the ANOVA of the non expected interaction yielded a null value of $\underline{F}$. In the opposite case, when subjects transferred from the two to the three-category task, overlap predicted an effect favoring categories $\underline{1}$ and $\underline{3}$ over $\underline{2}$, although this effect was predicted to be of less magnitude than that observed in Experiment 2. The results were in the direction that had been predicted, but they were only marginally significant.

Summing up, effects tended to show up when differences in overlap were large, were either not observed or only marginally so when differences in overlap were small, and were not observed at all when the overlap differences were null. At a general level then, this pattern of results is consistent with the hypothesis that overlap with respect to prior categories influences the process of constructing new categories. At a detailed level, the absolute predictive validity of mutual entropy as a measure of overlap remains to be determined.

The current work indicates that the process of building a new category can be considered as a particular case of knowledge generalization, and, as such, it is susceptible to be guided by whichever prior categories might be available to the learner. When a child learns that a particular dog is a mammal (a category label never encountered before), he might be able to generalize the "mammalness" property to all dogs. As a conclusion, prior categories will bias the process of forming new ones. In Keil's (1990) terms, prior categories constitute a domain-specific, acquired constraint on category formation.

Interestingly, the claim that overlap with prior categories facilitates category formation might not only contribute to the further understanding of the process of category formation, but might also provide a parsimonious causal explanation for the generally accepted claim that human categories tend to be organized hierarchically.

Acknowledgments

The work presented in this paper was conducted at the Georgia Institute of Technology as part of the author's doctoral dissertation, under the guidance of Professor Dorrit Billman, and the financial support of a MEC-Fulbright Scholarship from the Spanish Ministry of Education and the U.S. Information Agency.

References

Barsalou, L. W. (1983). Ad-hoc categories. *Memory and Cognition, 11*, 211-227.

Cabrera, A. (1995). The "rational" number *e*: A functional analysis of categorization. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence-Erlbaum.

Cabrera, Á., & Billman, D. (1996). Language-driven concept learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 539-555.

Corter, J. E. & Gluck, M. A. (1992). Explaining basic categories: feature predictability and information. *Psychological Bulletin, 111*, 291-303.

DeJong, G. & Mooney, R. (1986). Explanation based learning: An alternative view. *Machine Learning, 1*, 145-176.

Keil, F. C. (1983). On the emergence of semantic and conceptual distinctions. *Journal of Experimental Psychology: General, 112*, 357-385.

Keil, F. C. (1989). *Concepts, Kinds and Cognitive Development.* Cambridge, MA: MIT Press.

Keil, F. C. (1990). Constraints on constraints: surveying the epigenetic landscape. *Cognitive Science, 14*, 135-168.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89-115.

Mitchell, T., Keller, R. & Kedar-Cabelli, S. (1986). Explanation-based learning: A unifying view. *Machine Learning, 1*, 47-80.

Mooney, R. J. (1993). Induction over the unexplained: Using overly-general domain theories to aid concept learning. *Machine Learning, 10*, 79-110.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*, 416-432.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, MA: MIT Press.