The Effects of Category Generalizations and 
Instance Similarity on Schema Abstraction

Renée Elio and John R. Anderson 
Carnegie-Mellon University

Three experiments were designed to differentiate two models of schema abstraction. One model, called the generalization model, proposes that category generalizations, defined as feature combinations that occur frequently across study items, are abstracted during learning and used to classify transfer items. According to the other model, called the instance-only model, transfer items are classified according to their similarity to studied items. Study materials were constructed that either yielded category generalizations (generalize condition) or did not (control condition). Transfer items differed on whether they were classifiable by category generalizations and on their similarity to study items. In Experiments 1 and 3, accuracy and confidence on transfer items were better in the generalize condition than in the control condition. Experiment 2 manipulated the order in which generalizable study items were presented for study: Items were either blocked, so that those contributing to a category generalization occurred close in the study sequence, or randomly ordered. Study items were learned faster and transfer performance was better with blocked presentation than with random presentation. In all three experiments, there was an effect for the similarity of transfer items to study material. There was some evidence indicating better transfer performance on novel items that partially, rather than completely, fit a category generalization. The results support a schema abstraction model in which transfer is a function of similarity both to specific category instances and to higher order category information abstracted from those instances.

It is a ubiquitous phenomenon that people are able to detect regularities that characterize a category of stimuli simply from experience with category members. The success of this inductive process is not limited to well-defined categories; those for which a single rule or list of defining attributes will always predict category membership. For most real-world categories there may be several complex rules governing membership, none of which is singularly predictive. We will call the process by which people learn ill-defined categories from experience with exemplars schema abstraction. We differentiate this process from concept identification only because this term has traditionally denoted classification learning situations in which the categories are defined by a single rule, often derived through explicit hypothesis testing. Since the acquired infor-
An alternative theory of schema abstraction is the view that the abstracted category information is based on the frequency with which features and feature combinations occur across exemplars of a category (Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974; Reitman & Bower, 1973). We will refer to these as strength or frequency models. Hayes-Roth and Hayes-Roth (1977) proposed that the frequency of occurrence of all an exemplar's single features plus all possible combinations of these features (called property sets) make up the exemplar's representation. The frequency with which a property set occurred among all the encoded exemplars of a category determines its associative strength to that category. They propose that recognition of an exemplar is governed by the associative strengths of its property sets to the category or categories studied. The diagnosticity of a property set for a given category was defined as an increasing function of its associative strength to that category and a decreasing function of associative strength to the alternative categories. Stimulus sets can be created in which instances that are farther from the central tendency or prototype have higher property-set diagnosticity than instances that are closer to the prototype. Hayes-Roth and Hayes-Roth demonstrated with such material that property set diagnosticity, not prototypicity, predicted classification behavior.

As different as these models may seem, they share the assumption that some information, qualitatively different from the representation of individual instances, is abstracted, stored, and used in subsequent recognition and classification judgments. In contrast, Brooks (1978) and Medin and Schaffer (1978) have argued that a model positing only one level of information, instance information, can account for the previous schema abstraction results. In a series of experiments, Medin and Schaffer controlled the distance of transfer items to the prototypes of two categories while manipulating the similarity of the transfer items to individual category members. They demonstrated that the interitem similarity of training exemplars affected learning time and that subsequent recognition and classification ratings of new instances were a function
of their similarity to individual training ex-
emplars, not of their distance from category
prototypes. They proposed that classification
of a novel item is based on its similarity to
all the stored items; in other words, the prob-
ability it will be classified in Category 1 is
an increasing function of its similarity to all
the stored Category 1 items and a decreasing
function of its similarity to all the stored
Category 2 items. Medin and Schaffer of-
ered a simplified version of this assumption:
The item most similar to a novel instance is
retrieved and its category assignment is used
to classify the novel instance.
Medin and Schaffer noted that the method
of generating stimuli in most classification
learning experiments—creating category ex-
emplars by applying distortion or transforma-
tion rules to the category prototype—
causes the prototype to be the transfer item
most similar to members of its own category
and least similar to members of another cat-
egory. Thus, the similarity-to-stored-in-
stances model can account for superior per-
formance on prototypes and items close to
prototypes without positing an additional,
qualitatively different level of information.
It can also account for the result that pro-
totype classification suffers little with delay,
because even if some specific instances are
forgotten, other instances similar to the pro-
totype will remain. Hintzman and Ludlam
(1980) replicated this phenomenon with a
computer simulation that stores only ex-
emplar information and uses a best-match
rule in conjunction with a forgetting mech-
anism for classifying old instances and pro-
totypic instances at delay.
However, strength models of the feature-
set variety can account for the data offered
as evidence for similarity-to-stored-instances
models because they propose that individual
instances are augmented with, not replaced
by, higher order category information. Al-
though Medin and Schaffer effectively dem-
onstrated the inadequacy of a prototype
model (at least when small categories are
learned), their experiments were not de-
signed to contrast the assumptions of their
similarity-to-stored-instances model with the
assumptions of strength models. The purpose
of the present series of experiments is to dis-
tinguish between an instance-only model
proposing that transfer performance is a
function of similarity to stored exemplars
and a particular strength model proposing
that some higher level, qualitatively differ-
ent information is abstracted from, and re-
presented in addition to, specific instances and
used to classify new items.
The model we are contrasting with the
instance-only model is called the ACT gen-
eralization model, based on the ACT theory
(Anderson, 1976; Anderson, Kline & Bes-
ley, 1979), a model and a computer simu-
lation of declarative and procedural knowl-
edge. Using general learning mechanisms
and assumptions not designed specifically for
schema abstraction tasks, the ACT program
successfully replicated the recognition and
classification results of Franks and Brans-
ford (1971), Neumann (1974), Hayes-Roth
and Hayes-Roth (1977), and Medin and
Schaffer (1978), given their respective tasks
and stimuli (Anderson et al., 1979). How-
ever, having the generalization model ac-
count for these results simply contributes
another competing model to the already
large set of alternative schema abstraction
theories. We designed the present experi-
ments not simply to marshal support for the
ACT generalization model but to differen-
tiate the predictions of an instance-only
model and frequency-based strength models,
of which the ACT generalization model is
one version.

The ACT Generalization Model
A generalization is a pattern of frequently
c-occuring features in a set of data. Al-
though less specific than any pattern seen,
a generalization captures the regularities
across specific items. For example, we might
learn that one member of Club 1 is single,
Catholic, plays tennis, and works for the
government. We might subsequently learn
that a second Club 1 member is single, Pro-
estant, plays tennis, and works for the gov-
ernment. While we would store both these
specific feature patterns, the generalization
we would form that accommodates both
these specific descriptions of Club 1 mem-
bers would be Club 1 members are single,
play tennis, and work for the government.
Since religion differed in the specific de-
scriptions, it is not a part of the generalization about Club 1 members. Note that, in addition to the original study instances, the model proposes that only one feature set—the above generalization—is stored. ACT contrasts with the Hayes-Roth and Hayes-Roth (1977) model, which predicts all feature subsets will be stored. It should be obvious that, with only a moderate set of four or five feature stimuli, the number of possible feature combinations would be very large. Thus, one advantage of the generalization model is that it greatly reduces the amount of information that needs to be stored.

Each time a generalization successfully classifies a specific feature pattern, its representation in memory becomes stronger; for example, learning that another Club 1 member is Jewish, single, plays tennis and works for the government reinforces the Club 1 members are single, play tennis, work for the government generalization. According to the model, each time a pattern of features successfully classifies an item, not only is it strengthened but any patterns more general but still consistent with it is also strengthened. For example, the first description, single, Catholic, plays tennis, works for the government, could be classified on some later learning trial by matching the specific feature pattern one Club 1 member is single, Catholic, plays tennis, works for the government previously stored for this item. This specific pattern would be strengthened. In addition, the generalization consistent with this pattern, Club 1 members are single, play tennis, and work for the government, although not the basis for this particular classification, would also be strengthened. Over time, then, such a generalization will accrue more strength than any of the specific patterns that generated it. This greater memorial strength is reflected in the higher probability that a generalization rather than a specific instance will be accessed to classify instances. In other words, the above Club 1 member descriptions would eventually be categorized by matching the generalization about Club 1 members rather than by matching the specific patterns initially stored for them. Anderson et al. (1979) offer a more detailed description of the mechanisms we have outlined here.

The various schema abstraction results are easily accommodated by the generalization model. The more distant an item is from its category prototype, the less similar it is to the majority of items and the less likely it is to be classifiable by generalizations formed from more prototypic items. The generalization model can also account for the facilitative effect of high interitem similarity among training exemplars (Medin & Schaffer, 1978): If training items from different categories share a high degree of overlap, generalizations between them will not only compete for application but their strength will be decremented if they misclassify items during training. The model can also accommodate the Rosch and Mervis (1975) finding that an item’s classification and typicality ratings depend on its family resemblance, the degree to which it is similar to items within its category and dissimilar to items in alternative categories. The generalization model predicts that classification performance on transfer items equally similar to study items from alternative categories would be poor, again because generalizations from different categories would be equally likely to match such items.

The key factor that distinguishes the ACT generalization model from previous theories is that it is the first clear process model for representing and capturing the consequences of correlated attributes in the study exemplars. The basic prediction of the theory is that the greater the overlap among study items in feature sets, the better performance will be on test items that share this overlap. The experiments to follow test this basic prediction by manipulating both the amount of overlap of test items with the study sets and the ease with which subjects may notice this overlap.

Experiment 1

Our general plan for distinguishing the generalization model and an instance-only model was to manipulate the likelihood of forming category generalizations in two different sets of study exemplars while holding
the similarity of transfer items to the two study sets as constant as possible. In this way, any advantage for having studied the items that yielded generalizations would be attributed not to a higher degree of interitem similarity between those items and the transfer set but rather to the availability of generalizations. We also manipulated the type of transfer item. One type of transfer item could be classified by applying category generalizations if generalizations had in fact been formed from experience with generalizable study exemplars. The other type of transfer item was not classifiable by category generalizations. According to the ACT theory as developed by Anderson et al. (1979), a category generalization formed during study must completely match a transfer item in order to classify it. Given this “full match” view, the generalization Club 1 members are single, play tennis, and work for the government fully matches and would assign to Club 1 a transfer item such as single, Baptist, plays tennis, works for the government, but not a transfer item such as married, Baptist, plays tennis, and works for the government. Therefore, performance should be better on transfer items that match the category generalization than on transfer items that do not.

The manner in which this general design was realized in Experiment 1 can be illustrated best with a small portion of the experimental materials. Subjects read five-feature descriptions of people who belonged to either the “Dolphin Club” or the “Koala Club.” Subjects in the generalize condition studied descriptions such as

1. One member of the Dolphin Club is a Baptist, plays golf, works for the government, is college educated, and is single.

2. One member of the Dolphin Club is a Baptist, plays golf, works for a private firm, is college educated, and is married.

From these exemplars, we anticipated that they would form the generalization that a member of the Dolphin Club is a Baptist who plays golf and is college educated, since these are the features that these two club members have in common. After learning to classify items like 1 and 2 into the Dolphin Club (and other items into the Koala Club), subjects moved to a transfer task in which they were presented with new items like

3. This person is a Baptist who plays golf, is unemployed, is college educated, and is divorced.

4. This person is a Baptist who plays tennis, is unemployed, is college educated, and is divorced.

Description 3 is an instance of what we called a three-overlap transfer item. It overlaps with Study Items 1 and 2 on three features and, moreover, on the three features that form the generalization (Baptist, golf, college). Therefore, we would expect transfer performance on Item 3 to be quite high, since the generalization formed from 1 and 2 matches it completely. In contrast, Description 4 only overlaps with the original study items on the two features Baptist, college. While both of these features are part of the generalization, we would expect a lower probability of classifying this item as a Dolphin Club member, since a generalization must match an item perfectly to be used. Therefore, a two-feature overlap with a generalization that requires three features should not help.

The other study condition was called the control condition. Rather than studying a pair of items like 1 and 2 above, subjects might study

5. One member of the Dolphin Club is a Baptist who plays golf, works for a private firm, is high school educated, and is divorced.

6. One member of the Dolphin Club is a Baptist who plays golf, works for the government, is college educated, and is married.

Note that these study pairs only overlap on two features, Baptist and golf. After learning to classify these items, subjects in the control condition were asked to judge the same transfer items as subjects in the generalize condition. Note that Transfer Item 3 is still a three-overlap item for subjects who have studied Items 5 and 6. It overlaps with 5 and 6 on three features, but a different set of three features for each study item (with 5 on Baptist, golf, divorced, and with 6 on Baptist, golf, college). According to the view that interitem similarity governs classification judgment, performance on 3 should
not differ depending on whether subjects studied Items 1 and 2 or 5 and 6, since the overlap of transfer item with study items is the same. However, the generalization point of view predicts an advantage for having studied Items 1 and 2, which offered a three-feature generalization for classifying 3, over studying Items 5 and 6, which only offered a two-feature generalization. The ACT generalization theory predicts poorer performance having studied 5 and 6, because the probability of forming a generalization, the probability of its applying to a test stimulus if formed, and the confidence that the subject will have in its application all increase with the number of features in the generalization. Note also that Transfer Item 4 overlaps with Items 5 and 6 on two features (but a different set of two features for each). Thus Item 4 is a two-overlap item in the control condition as well as in the generalize condition.

The other factor manipulated during learning was whether subjects saw pairs like Items 1 and 2 or like 5 and 6 close together in the study sequence of study items or randomly spread apart. This was the blocked versus random presentation manipulation. We expected the transfer performance of generalization subjects to be better in the blocked condition than in the random condition: If generalizable pairs are close together, they are more likely to be simultaneously available in a working memory for patterns. In contrast, we did not predict any particular difference between blocked and random conditions with control study materials.

To summarize, the ACT generalization model predicted performance to be best in the generalize-blocked condition on three-overlap transfer items, since these items are classifiable by category generalizations, and equally poor on all other types of transfer items. An instance-only model predicts no effects of generalize versus control study material, nor does it predict an advantage for blocking. Although it would predict an advantage for three-overlap versus two-overlap transfer items, it does not predict that this effect would vary with study material (generalize vs. control) or blocking. In contrast, the generalization model predicts an interaction of study material with transfer item type, with the largest effect of three-overlap versus two-overlap transfer items for subjects in the generalize-blocked condition.

**Method**

**Subjects.** Eighty members of the Carnegie-Mellon University community served as subjects. They received 2 credits per week and/or $3 an hour for their participation. Twenty subjects were exposed in each condition: generalize-blocked, generalize-random, control-blocked, control-random. Subjects were randomly assigned to one of the four experimental conditions. The experimental session lasted approximately 2 hr.

**Material and design.** The stimuli were five-feature descriptions of people to be classified members of one of two clubs. Each feature had four possible values. The five features and their values were: job—(1) unemployed, (2) self-employed, (3) government, (4) private firm; marital status—(1) single, (2) married, (3) divorced, (4) widowed; religion—(1) Catholic, (2) Jewish, (3) Episcopalian, (4) Baptist; hobby—(1) tennis, (2) golf, (3) chess, (4) bowling, education— (1) grammar school, (2) high school, (3) college, (4) trade school. Each stimulus item could be described symbolically as five digits, one for each feature, with each digit ranging from 1 to 4 to indicate the specific value of each feature. Given the above assignment of digits to values, for example, the item 43211 could correspond to the description private firm, divorced, Jewish, tennis, grammar school. The design of the study and transfer material, given in Table 1, was specified by these numbers rather than the specific feature values. The values of each feature were randomly reassigned to the digits 1-4 independently for each subject. The order of the features in the description was also randomly determined for each subject. This means that for one subject, the 11114 item from Table 1 might have meant government, single, Baptist, high school, chess, whereas for another subject it might have meant Baptist, golf, private firm, college, married. Thus, each subject had his or her own randomly generated set of materials.

Table 1 schematically illustrates the stimuli and design of the experiment. A 2 (study set) x 3 (test item type—the two transfer item types plus the originally studied items) x 2 (presentation order) design was used. Study set and presentation were varied between subjects. Table 1 shows items from the two study sets, the generalize set and the control set. Pairs of items in the generalize set gave rise to three-feature generalizations. The four Club 1 generalizations were 11-1, 1-22, 4-11, and 22-4. The pairs of study items in the control condition shared only two features. For both study sets there was no value on any feature that could perfectly predict club membership. The values 1 and 2 predicted Club 1, since they occurred more frequently on each feature than the values 3 and 4. Club 2 items were constructed by interchanging 1s and 4s with each other and 2s and 3s with each other, so that the values 3 and 4 predicted Club 2. The third feature was irrelevant with respect to club membership.

Since the critical aspects of the design rest on the
relationships between the study item sets and the test items, it is worthwhile to work through an example. The two study sets were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 11122. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items.

To help explain the various transfer conditions, let us go through an example of each transfer item type in Table 1. Consider the first Club 1 generalize study pair, 11114 and 1112. Two items were constructed so that pairs of items in each set were equated for the amount of overlap they had with pairs of transfer items. The two types of transfer items were defined with respect to a study pair. Transfer items can be classified according to their overlap relationship to their corresponding study pair. Overlap was the number of features for which two items had the same values. Transfer items could overlap on either three features (three-overlap) or two features (two-overlap) with their corresponding study pair. For example, the item 14322 has a three-feature overlap (three-overlap) with the item 12122 on the first, fourth, and fifth features, since the values 1, 2, and 2, respectively, are the same for both items. The overlap characteristics of the transfer items are also true with respect to the control study set. For example, the three-overlap transfer items, 11313 and 11413, overlap the first control study pair, 11114 and 1112, on the first, second, and fourth features and with the second control study items, 11223, on the first, second, and fifth features. Thus, these two transfer items have in common the features with both items in their respective control study pair as they did with both items in their corresponding generalize study pair. The critical difference is that the three-overlap with the generalize study pair matched a category generalization, whereas the three-overlap with the control study pair did not. A comparison of this control study pair with other transfer items will indicate that the same relations described above for the generalize study set held for the control study set.

There were 16 items in each study set and in each transfer item type, half Club 1 members and half Club 2 members. For blocked presentation, the 16 study items were divided into four groups of four items each. Each group consisted of one Club 1 study pair and one Club 2 study pair. The order of presentation within each group of four items was permuted and the final sequencing of the four groups for study was randomly determined. This method assured that, in the generalize condition, generalizable items were separated by at least two intervening items. Random presentation was realized as a pseudorandom ordering of the items using a method similar to the one described above for blocking. The difference was that the Club 1 items and the Club 2 items combined into a group of four were selected from different study pairs. The actual ordering of the items differed from trial to trial within the constraints of the blocking and random ordering algorithms.

Apparatus. The experiment was controlled by a PDP 11/34 computer. Subjects were seated in three adjoining three-person rooms, each of which contained a CRT screen on which the stimuli were displayed.

Procedure. The experiment was divided into two phases, a study phase and a test phase. For the study phase, subjects were told that their task was to learn to classify 16 people as either Dolphin Club or Koala Club members on the basis of their description (club names were chosen to correspond to the terminal response keys "d" and "k"). To encourage subjects to attend to all five features, they were told that club membership was determined in a complex fashion and that there was no bias with respect to membership on the basis of a single feature. They were also encouraged not to formulate and test hypotheses during learning, but to concentrate on memorizing each description with its club assignment. In all the experiments reported here this instruction was used to induce as little analytic processing as possible. This was desirable in light of Brooks' (1975) distinction between analytic and nonanalytic processing concerning the nature of the information encoded about a stimulus item. Specifically, Brooks suggests that a learner in an analytic mode may not encode all aspects of the stimulus, but only those dimensions and/or values that seem relevant to the current hypothesis. It is likely that our subjects would spontaneously generate and test hypotheses corresponding to the complex rules that govern category membership in these experiments. It is important to note the distinction between conscious rule generation and the automatic generalization mech-
anation that is part of the ACT model. In order for the generalization mechanism to be as successful as possible in detecting feature patterns, the data base on which it operates must be as veridical as possible. Hence we stressed item memorization rather than active hypothesis testing. The stimuli were presented in blocks of 16. One pass through all 16 items constituted one trial. The learning criterion was set at one correct pass through all the 16 items; that is, one 100% accurate trial. Subjects kept cycling through the 16 items until they reached this criterion. The study items were presented one at a time in the middle of the terminal screen. Subjects hit either "d" or "k" to classify a person as a Dolphin Club member or a Koala Club member, respectively. If a response was entered, feedback of the form "Right, Dolphin (Koala) Club" or "Wrong, Dolphin (Koala) Club" appeared on the screen. The description, the subject's response, and the feedback remained on the screen for 10 sec. The screen then erased and the next item was presented. A 10-sec response-time limit was set. If the subject did not classify the item within 10 sec, the correct club membership appeared, followed by the 10-sec study time. Subjects were informed that failure to respond within 10 sec counted as an error. At the end of each pass through the 16 items, subjects were told their accuracy for that trial. There were rest breaks after every fourth trial.

After reaching the learning criterion, subjects began the test phase. They were told their task was to classify a new set of people as quickly as possible without sacrificing accuracy. Both the study items and the transfer items were presented during the test phase in a different random order for each subject. The test items were presented one at a time in the center of the screen and subjects hit either "d" or "k" to classify the description. After the subject classified the item, the word confidence appeared on the screen. Subjects assigned a confidence rating to their judgment, ranging from 1 (not at all confident) to 5 (absolutely confident). Subjects were informed that the confidence rating was not timed and were encouraged to make sure it accurately reflected how confident they felt about their judgment. The description and the subject's response remained on the screen until the confidence rating was made. The screen then erased and the next item was presented. Accuracy and confidence ratings were recorded for each classification.

Results

The mean number of trials to criterion in the study phase was 12.05 for generalize-blocked, 13.50 for generalize-random, 16.95 for control-blocked, and 15.85 for control-random. The effect of study set was significant, $F(1, 76) = 6.3, p = .014$. Newman-Keuls tests indicated that the generalize-blocked and the control-blocked conditions differed significantly, but the difference between the study set conditions with random presentation was not significant by this test. Although learning in the generalize-blocked condition was faster than in the generalize-random condition in the predicted direction, neither the blocking manipulation nor its interaction with study set was significant, $F(1, 76) = 1.0$. Since both study set tests had equivalent ratios of diagnostic to nondiagnostic values on each feature, faster learning in the generalize conditions could not be attributed to the use of independent, diagnostic cues.

Confidence scores were computed as the mean of a subject's confidence ratings on correct classifications minus his or her confidence ratings on incorrect classifications for a given test item type. Thus, confidence scores range from −5 to +5. Accuracy on the study items at retset for generalize-blocked was 88%, for generalize-random was 85%, for control-blocked was 81%, and for control-random was 79%. The mean confidence ratings of study items in these four conditions were 3.16, 3.38, 2.73, and 2.84, respectively. Although suggestive, the variation among these conditions on accuracy and confidence was not significant. The less than perfect performance on study items after reaching criterion during study probably reflects both successful guessing to reach study criterion and the subject's forgetting of his or her decision rules in the face of interfering transfer items.

Table 2 presents the mean accuracy and confidence rating for each item type within each condition. Analyses of both the accuracy and confidence data for transfer items revealed a significant advantage for the generalize condition over the control condition, $F(1, 76) = 15.4$ and $14.3$, respectively, $p < .001$. For each transfer item type, subjects in the generalize condition were more accurate and more confident than subjects in the control conditions. There was a significant effect of type of transfer item on accuracy, $F(1, 76) = 4.2, p < .001$, and confidence, $F(3, 76) = 3.3, p < .001$. Newman-Keuls tests on both the accuracy and confidence means revealed that all pairwise comparisons of item types differed significantly.

1 All Newman-Keuls tests reported were significant at the .05 level.
GENERALIZATIONS AND SCHEMA ABSTRACTION

Table 2: Mean Accuracy and Confidence Scores on Transfer Items as a Function of Study Material and Presentation Order, Experiment 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Generalize</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blocked</td>
<td>Random</td>
</tr>
<tr>
<td>Three-overlap</td>
<td>79</td>
<td>77</td>
</tr>
<tr>
<td>Two-overlap</td>
<td>70</td>
<td>71</td>
</tr>
<tr>
<td>M</td>
<td>75</td>
<td>74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Confidence</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-overlap</td>
<td>2.43</td>
<td>2.53</td>
<td>1.49</td>
<td>1.83</td>
<td>2.07</td>
</tr>
<tr>
<td>Two-overlap</td>
<td>1.66</td>
<td>1.77</td>
<td>0.57</td>
<td>1.37</td>
<td>1.34</td>
</tr>
<tr>
<td>M</td>
<td>2.05</td>
<td>2.15</td>
<td>1.03</td>
<td>1.60</td>
<td>1.71</td>
</tr>
</tbody>
</table>

* Percent correct.

...cantly. The blocking manipulation had no appreciable effect on either accuracy or confidence, nor did it enter into any significant interactions.

Discussion

The results in Experiment 1 indicate that transfer to new category exemplars is facilitated when studied exemplars yield generalizations. In addition, initial learning of the study items was facilitated when generalizations existed between items being learned.

An instance-only model cannot account for the beneficial effects of generalizations on transfer performance. However, the generalization theory is not unequivocally supported. For both the generalize and control conditions, classification performance was also a function of similarity to studied exemplars: The less similar transfer items were to studied items, the worse classification performance was. Under the view that a generalization must match a test item perfectly to apply, the generalization theory would predict good performance on the three-overlap transfer items, to which the category generalizations apply, and chance performance on the two-overlap items. Yet subjects performed well above chance in classifying transfer items for which generalizations did not completely match (two-overlap items) and transfer items for which generalizations did not exist at all (control condition). Thus, we cannot exclude the importance of similarity to studied items in this task.

A careful postexperimental examination of the stimuli uncovered some unintended variation. Although the test items had satisfied our overlap constraints with the intended study pairs, they had a number of spurious overlaps with other study pairs. For example, although a two-overlap item did in fact have only two features in common with each of its corresponding study items, it may have overlapped on three features with some other study items. To assess the extent of these spurious overlaps, we computed an overlap score for each test item to the generalize set and to the control set in the following manner. Each transfer item had two overlap scores. Its positive overlap score represented how similar it was with study items in its assigned category. Its negative overlap score represented how similar it was with study items in the alternative category. For each transfer item, we tabulated the frequency of five, four, three, two, and one overlaps it shared with all the study items in its assigned club (e.g., each Club 1 transfer was compared with all the Club 1 study items). Using a metric similar to the one advocated by Medin and Schaffer (1978), these frequencies were weighted by the square of the amount of overlap they represented (e.g., the number of three overlaps was weighted by nine, the number of two overlaps by four,
and so on) and summed. This was the transfer item's positive overlap score, its similarity to study items in the category to which it was assigned. A transfer item's negative overlap score was computed in the same way, except that the transfer item was compared with study items in the alternative club (e.g., each Club 1 transfer was compared with all the Club 2 study items). A transfer item's final overlap score was the difference between its positive and negative overlap scores. Each transfer item had an overlap score for both the genericize study materials and the control study materials.3

Table 3 gives the mean overlap score for each transfer item type with the generalize study set and with the control study set. The means for the study items represent their interitem similarity. Note that, using this metric, the control study items, compared with the genericize items, had the same or more interitem similarity. Given these equivalent interitem similarity scores for the two types of study materials, an instance-only model would be at a loss to explain the significantly faster learning of genericize study items. On the other hand, there is greater similarity for the transfer items in the generalize condition than in the control conditions. There are highly significant correlations (p < .001) between accuracy, confidence, and similarity for transfer items in both the control and the generalize condition.

Table A1 in the Appendix presents the performance on individual items in the transfer test. The design of these experiments was such that for each Club 1 item there was a Club 2 item that was identical in all respects (indeed, Club 1 and Club 2 could be switched by simple redefining the digits 1–4 in Table 1). In Table A1, we present the data averaged over these corresponding pairs of Club 1 and Club 2 items and refer to this average by the Club 1 numerical notation. Relatively small numbers of observations contributed to these individual items, and performance measures for these individual items are not particularly stable. Still, some theory might regard a portion of the variance among individual items as systematic. Note, for example, the difference in accuracy and confidence between Items 14322 and 41111.

### Table 3

<table>
<thead>
<tr>
<th>Study condition</th>
<th>Test item</th>
<th>Generalize</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study overlap</td>
<td>22.0</td>
<td>23.5</td>
<td></td>
</tr>
<tr>
<td>Three-overlap</td>
<td>17.2</td>
<td>15.3</td>
<td></td>
</tr>
<tr>
<td>Two-overlap</td>
<td>10.1</td>
<td>2.7</td>
<td></td>
</tr>
</tbody>
</table>

*This score is a measure of the amount of interitem similarity among the study items themselves.

Both these items have a similarity score of 17 in the generalize condition, but the accuracy and confidence means for Item 14322 are considerably higher than those for Item 41111. Item 14322 can also be classified by virtue of the value 2 on its fourth dimension; that is, Value 2 on this dimension never occurs in Club 2 items (and conversely, Value 3 on this dimension never occurs in Club 1 items). Item 14422 can also be classified by this rule. Similarly, Value 2 on Dimension 1 is a sufficient predictor of Club 1 membership for items 22142 and 22242. The accuracy and confidence means in Table A1 for these sufficient-feature items are generally higher than for other items with comparable similarity scores. A theory in which the search for sufficient features or feature combinations drives a generalization mechanism may be able to account for this variance in the data.3 We will return to this distinction between sufficient-feature items and no-sufficient-feature items later in the article.

We looked at performance on specific transfer items to determine if similarity to studied items alone could account for our results. We paired three-overlap transfer items from the control and generalize conditions such that the control item had as high

---

2 There are other ways of calculating similarity or overlap. We have tried a couple of others, and they yield substantially the same conclusions. We chose to use a metric based on the work of Medin and Schaffer because theirs is the most successful instance-based model.

3 The authors wish to acknowledge Doug Medin for bringing the sufficient-feature aspect of the stimulus set and its implication to our attention.
These item analyses bolster our claim that transfer performance in Experiment 1 cannot be accounted for solely by a similarity-to-stored instances model. We also acknowledge the inadequacy of the generalization model, in its present form, in accounting for the similarity effects we found. We will elaborate on the inadequacy of the generalization model in the face of these similarity effects and propose certain reformulations of the model in the General Discussion. The next experiment focused on the effect of blocked versus random presentation, and the third experiment focused on the effect of generalize versus control study materials. By examining each of these factors one at a time, we were able to avoid the design constraints that led to the large amount of uncontrolled variation in overlap between study items and transfer items in Experiment 1. In addition, by focusing on a single factor at a time, we were able to perform a more powerful manipulation of each variable and also get the added statistical power of a within-subjects design.

**Experiment 2**

In Experiment 2, we contrasted two generalize conditions, one in which forming generalizations might be facilitated by blocking and one in which forming generalizations was hindered by random presentation of instances. To enhance the effect of blocking, we increased the ratio of items to generalizations, so that a given generalization accounted for three exemplars per category rather than just two. The strength of the resulting generalizations should be greater than in the previous experiment and the potential for blocking to have an effect should be greater. To increase the statistical power of the experiment, we made the presentation-order manipulation a within-subjects factor by running a two-phase experiment. In Phase 1, subjects studied generalizable items presented either blocked or randomly. In Phase 2, subjects studied a different set of generalizable materials in the alternative presentation order.

**Method**

**Subjects.** Forty-three members of the Carnegie-Mellon University community received psychology course
credit and/or $3 an hour for their participation in the 2-hr. experiment.

Materials and design. Two sets of stimulus items were used to create five-feature descriptions of people to be classified as Dolphin or Koala Club members and of "space invaders" to be classified as "friendly" or "hostile." Each of the five features made up the generalizations would be different in each of the two phases of the experiment so that the form of the generalizations (i.e., which of the five features made up the generalizations) would be different in each of the two phases (see Table 4 for Stimulus Set A and Table A2 in the Appendix for Stimulus Set B).

Each feature had six values. The club member features and values were: job—unemployed, self-employed, government, private firm, military, retired; religion—Catholic, Jewish, Episcopalian, Baptist, Mormon, Lutheran; hobby—stamps, coins, painting, gardening, chess, reading; musical taste—classical, jazz, rock, disco, folk, country; sport—volleyball, basketball, bowling, squash, racquetball, handball. The space invader features and values were: color—purple, red, blue, green, yellow, brown, skin—metallic, furry, spiny, scaly, translucent, luminescent; appendage—claws, antennae, horns, wings, tentacles, tail; home planet atmosphere—radon, neon, helium, xenon, argon, krypton; base of operations—Venus, Mars, Jupiter, Saturn, Uranus, Pluto. As in Experiment 1, the ordering of features in the description and the assignment of descriptive values to numeric values were randomly determined for each subject. There were nine items in each category. Study items were generated in sets of three. The three items in a set shared three features in common. Thus, there were three generalizations per category. The three Category 1 generalizations were 11111-, 11114-, and 11121-. The corresponding Category 2 generalizations were 4444-, 4441-, and 4442-.

Three types of transfer items were constructed: four-overlap, three-overlap, and two-overlap. The four-overlap items shared the generalization yielded by one of the study set triplets plus a fourth feature with some of the items in the triplet. For example, the study item triplet 11144, 11121, and 11132 yields the generalization 111-. The four-overlap item 11134 shares the 111—generalization with each of these three study items. It also overlaps the first study item on the fifth feature and the third item on the fourth feature. There were 18 (9 per category) four-overlap transfer items.

The two-overlap transfers overlapped on only two features with any study item in their respective category. A computer program generated all possible two-overlap items for each stimulus set. From this set, we selected the items that had a relatively high (four or more) number of positive two-overlaps and a relatively low (two or fewer) number of negative two-overlaps. Some of these two-overlap transfers had the property that the two features they shared with a study item matched part of a category generalization. For example, the three study items 11114, 11112, and 11132 have the generalization 111—, and the two-overlap transfer item 12113 overlaps on the first and third features of each of these items and with the generalization as well. Two-overlap items that had this property were designated as two-overlap partial matches—two (PM) overlaps—since they matched two thirds of a category generalization. In contrast, a two-overlap item such as 12214 also overlapped three study items (11114, 14312, and 32232) on two features, but none of these two-feature overlaps partially matched any of the category generalizations. For Stimulus set A, 6 out of 18 two-overlap transfers were partial matches. For Stimulus set B, 10 out of 18 two-overlap transfers were partial matches. The partial matches are starred in Table 4 and in Table A2.

A third type of transfer item, three-overlap transfers, was also included. These items matched one of the three-feature category generalizations yielded by one study item triplet. However, they were qualitatively different from the other transfer items, since one of their non-overlapping features had values that were not used in any of the study items. In other words, the study items used only four of the six possible values on a given fea-

<table>
<thead>
<tr>
<th>Study items</th>
<th>Transfer items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>Category 2</td>
</tr>
<tr>
<td>11132</td>
<td>44423</td>
</tr>
<tr>
<td>11144</td>
<td>44813</td>
</tr>
<tr>
<td>11121</td>
<td>44434</td>
</tr>
<tr>
<td>14312</td>
<td>41243</td>
</tr>
<tr>
<td>24412</td>
<td>31143</td>
</tr>
<tr>
<td>44112</td>
<td>11443</td>
</tr>
<tr>
<td>32223</td>
<td>23332</td>
</tr>
<tr>
<td>42323</td>
<td>12332</td>
</tr>
<tr>
<td>21223</td>
<td>43332</td>
</tr>
</tbody>
</table>

* Asterisks indicate items that partially match a category generalization.
tature, but the remaining two values were used to construct the three-overlap transfer items. For example, the three-overlap item 11115 overlaps each of the study items 11144, 11121, and 11132 of the 111...generalization, but the value 6 on the fifth feature was never used in any study item. It was necessary to use new values in order to construct items that overlapped on only the three-feature generalization for one category but did not overlap on three features with an item in the alternative category. Since three-overlap transfers contained never-studied values, they were always presented as the last items in the transfer test. This was done to ensure that performance on four- and two-overlap transfers was uncontrolled by any "surprise" effects these items might generate. There were 12 (6 per category) three-overlap transfers in each stimulus set. To summarize, category generalizations could be used to classify four-overlap and three-overlap transfer items, according to the full-match view, they would not be helpful in classifying either two-overlap or two(PM)-overlap transfer items.

The presentation factor (blocked vs. random) varied within subjects. In one phase, a subject's study items were blocked, and in the other, the study items were presented randomly. For blocked presentation, two study item triplets, one from Category 1 and one from Category 2, were randomly selected to be combined as a group of six items, whose order was then permuted. A second pair of study item triplets was selected, combined, and permuted as a group of six. The final pair of triplets was then permuted. These 18 study items were presented on one trial in this order. This method assured that the three items yielding a given generalization were clustered relatively close in the presentation sequence. For random presentation, items were also sorted into three groups of six, but the items in a given group of six came from each of the six different triplets. None of the three Category 1 items in a group of six were from the same Category 1 triplet, so they did not yield any category generalizations among themselves. Thus, there were no generalizable pairs in any block of six items. The order of the six items in each group was permuted, and the 18 items were presented in this order. The actual ordering of items in the blocked and random conditions varied from trial to trial, within the constraints of their respective presentation algorithms.

Apparatus and procedure. The apparatus and procedures were the same as described for Experiment 2. The only difference was the learning criterion. To assure that subjects would complete both phases of the experiment in the allotted time, the learning criterion was set at 85% correct passes through the 18 study items. If a subject did not reach this criterion after 14 passes, she or he moved on to the transfer test of the phase.

Results

There were 14 cases of failure to reach the learning criterion. Six of these were three subjects who did not reach criterion in either phase of the experiment. Of the remaining eight cases, five occurred with blocked presentation and three occurred with random presentation. The mean number of trials to criterion for learning was 8.2 in the blocked phase and 9.5 in the random phase (based on all subjects, including those failing to reach the learning criterion in 14 trials). The advantage of blocked presentation was significant, F(1, 41) = 4.4, p = .041. When the order of phases, blocked-random or random-blocked, is considered as a between-subjects factor, blocked-random subjects averaged 8.9 trials on their blocked (first) phase and 9.2 trials on their random (second) phase. Random-blocked subjects averaged 9.7 trials in their random (first) phase and 7.5 trials in their blocked (second) phase. Although the trends are suggestive, the effect of phase order was not significant, nor was the interaction with presentation. Subjects always learned faster in their blocked phases than in their random phases.

In the blocked condition, the mean accuracy on study items at retest was 79% and the mean confidence was 2.58. For random phases, the scores were 79% and 2.56. These retest scores were about 7% lower than those found in Experiment 1. However, none of the Experiment 2 items had the sufficient-feature characteristic, which may have boosted subjects' initial learning and retest performance on study items in Experiment 1. The mean accuracy and confidence scores for transfer items are presented in Table 5.

Analyses of variance on the transfer data included the order of blocked and random phases as a between-subjects factor. There were 21 subjects whose phase order was blocked-random and 22 subjects whose phase order was random-blocked. The two-overlap transfer items were partitioned into two(PM) overlaps and two overlaps, making a total of four transfer item types. There was a main effect of presentation on accuracy, F(1, 41) = 5.2, p = .029. Subjects' mean accuracy on transfer items was 73% in blocked phases but only 67% in random phases. The means in Table 5 show that accuracy varied greatly as a function of transfer item type, an effect that was highly significant, F(3, 123) = 31.5, p < 001. Not surprisingly, subjects were most accurate on four-overlap transfers and least accurate on two-overlap
transfers. Newman-Keuls tests on the blocked-phase accuracy means indicated that accuracy on each of the two types of high (four and three) overlap transfers was significantly higher than accuracy on each of the two low (two and two(PM)) overlap transfers. Similarly, random-phase accuracy means for four- and three-overlap items were significantly higher than accuracy on all the two-overlap items. It is interesting to note that subjects’ accuracy on two(PM)-overlap items, which partially matched two of the three features of a category generalization, was significantly higher than their accuracy on two-overlap items. Apparently, having a partial overlap with the generalization led to an advantage.

Subjects were also more confident in blocked presentation conditions, but unlike the accuracy results, this effect did not reach statistical significance, $F(1,41) = 2.8, p = .1$. There was a significant effect of transfer item type on confidence scores, $F(3,123) = 39.1, p < .001$. Newman-Keuls tests revealed that all pairwise comparisons of confidence means for the transfer item types were significant, with the exception of the three-overlap and two(PM)-overlap contrast.

**Discussion**

Support for the proposal that generalizations are formed during learning and used during transfer comes from a number of sources in Experiment 2. First, learning was faster when generalizable items were blocked than when they were randomly ordered. Second, transfer performance was higher in the blocked phases than in the random phases. An exemplar-based model in which similarity determines the acquisition and representation of the initial study items may be able to account for these findings. Finally, there was no interaction of presentation mode with test item type: The effect of decreasing similarity of test items to study items was the same in both blocked and random conditions. While blocking items may have facilitated forming generalizations, the effect of transfer item type in the random condition suggests that some category generalizations were formed even when generalizable items were randomly ordered. The third piece of (unexpected) evidence that argues for the existence of generalizations is the better transfer to two(PM)-overlap items relative to two-overlap items. Subjects were more accurate on two-overlap transfers that partially matched a generalization than on those that did not, a result the generalization model would not have predicted.

To better evaluate this advantage for partial matches to generalizations, we computed an overlap score for each of the two(PM)-overlap and two-overlap items, using the method described in Experiment 1. Although each two-overlap and two(PM)-overlap transfer had only two and one overlap with the study items, this similarity measure is based on the frequency with which the overlaps occur. The mean overlap score for the two(PM)-overlap items, set A and B combined, was 9.25. For the two-overlap items, the mean overlap score was 8.0. Although the differences between the two(PM)-overlap items and the two-overlap items on these two measures are small, they could account for the performance differences we found. We performed an analysis of covariance using item type, two(PM)-overlap versus two-overlap, as the random factor and individual transfer item overlap scores as the covariate. With this analysis, the effect of item type on accuracy approached significance, $F(1,33) = 3.6, p = .06$. This analysis suggests that the two(PM)-overlap advantage did not occur simply because these items had more two-feature matches with study instances than the two-overlap items did, but because some of those two-feature matches also partially matched a generalization. In other words, there seemed to be an effect for similarity to feature patterns that were category generalizations. However, although the overall performance differences on low overlap items that do and do not partially match category generalizations are suggestive, the role of partial matches to generalizations in classification judgments warrants a more controlled investigation of its own.

**Experiment 3**

Experiment 3 was designed as an attempt to replicate the Experiment 1 result that transfer to new items was better if studied items yielded category generalizations than if they did not. As in Experiment 1, we con-
<table>
<thead>
<tr>
<th>Presentation order</th>
<th>Transfer item</th>
<th>Blocked</th>
<th>Random</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy*</td>
<td>Four-overlap</td>
<td>83</td>
<td>78</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Three-overlap</td>
<td>78</td>
<td>72</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Two(PM)-overlap</td>
<td>69</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>73</td>
<td>67</td>
<td>70</td>
</tr>
<tr>
<td>Confidence</td>
<td>Four-overlap</td>
<td>2.80</td>
<td>2.42</td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td>Three-overlap</td>
<td>1.86</td>
<td>1.63</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>Two(PM)-overlap</td>
<td>1.24</td>
<td>1.03</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Two-overlap</td>
<td>0.87</td>
<td>0.36</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>1.69</td>
<td>1.36</td>
<td>1.53</td>
</tr>
</tbody>
</table>

* Percent correct.

trasted a generalize study set whose items yielded three-feature generalizations with a control set whose items did not yield cate-
gory generalizations. In Experiment 1 there turned out to be some differences between the two sets of material in the similarity be-
tween study and transfer items. Although there was still an advantage for the generalize condition when we looked at subsets of items for which this similarity difference was not a problem, it would be desirable to show a generalize advantage for having studied generalizable materials that were more equiva-
mental to control materials overall in terms of interitem similarity of study and transfer items. We discovered that we could not generate control and generalize study material that were equally similar to the transfer items and that satisfied overlap and cue validity constraints if we used the same transfer item set for both generalize and con-
trol study sets as in Experiment 1. Instead, we designed a generalize study set and a co-
rol study set each with its own transfer item set and tried to make the relation be-
tween the transfer set and the study set as equivalent as possible for both the generalize and the control materials.

Method

Subjects. Forty members of the Carnegie-Mellon University community received psychology course credit and/or $3 an hour for participation in the 2 hr. experi-
ment.

Materials and design. The club member and space in-
vader materials described for Experiment 2 were used in conjunction with the generalize and control items presented in Table 6.

For both the generalize and control study items, there were four possible values for each of the first four features and five possible values for the fifth feature. Cat-
gory 2 was constructed from category 1 by interchanging 1a with 4a and 2a with 5a. The value 4 on any feature was predictive of Category 1 and the value 4 on any feature was predictive of Category 2. Except for the fifth feature, the value 2 was somewhat predictive of Category 1 and the value 3, of Category 2. The values 2, 3, and 5 on the fifth feature were not predictive of either category.

There were eight study items per category. For the generalize condition, pairs of study items were con-
structed to overlap on three features. The four Category 1 generalizations were: 112,-- 121,-- 211,-- and 112,--. For the condition, study items were also constructed in a pairwise manner. The items in a control study pair shared only one feature in common. Using the metric described in Experiment 1, the mean overlap score for the generalize study items, a measure of their interitem similarity, was .28. For control study items, the mean overlap score was 25.1.

Only one type of transfer item, three-overlap, was used; there were eight three-overlap transfer items per category. One pair of transfer items was constructed for each pair of study items. For the generalize study set, each of the transfer items in a pair overlapped each of the items in its corresponding study pair on three features. These three features were the generalization yielded by the study pair. For example, the study pair 11225 and 11224 overlapped on the first, second, and third features, yielding the generalization 112,--. The two transfer items 11213 and 11224 overlapped each of the study items on the first three features and are classified by the 112,-- generalization. A transfer item pair overlapped on three features only with the items in its cor-
responding study pair; that is, there were no unexpected threecfeature overlaps between a transfer item and a third study item.

Pairs of transfer items were constructed in the same way for the control study set. Each transfer item in a pair overlapped with two study items on three features in its corresponding study pair, but a different three features for each item. For the control study pair 11235 and 1241, the transfer item 11241 overlaps the first study item on the first, second, and third features and overlaps the second study item on the first, fourth, and fifth features. The second transfer item 12135 overlaps the first study item on the first, fourth, and fifth features and the second item on the first, second, and third items. As in the generalize study set, a control transfer pair shared three features only with its corresponding study pair.

An overlap score was computed for each transfer item by tabulating the frequency of three, two, and one overl-
laps with study items in its assigned category (positive overlap) and with study items in the alternative category (negative overlap). As in Experiment 1, we multiplied the frequency of each overlap type by the square of the overlap and summed the results. The positive overlap
Table 6

<table>
<thead>
<tr>
<th>Study Items and Transfer Items for Generalize and Control Conditions, Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalize</td>
</tr>
<tr>
<td>Category 1</td>
</tr>
<tr>
<td>Study items</td>
</tr>
<tr>
<td>11213</td>
</tr>
<tr>
<td>11241</td>
</tr>
<tr>
<td>12311</td>
</tr>
<tr>
<td>12414</td>
</tr>
<tr>
<td>23113</td>
</tr>
<tr>
<td>24111</td>
</tr>
<tr>
<td>31122</td>
</tr>
<tr>
<td>41121</td>
</tr>
<tr>
<td>Transfer items</td>
</tr>
<tr>
<td>11213</td>
</tr>
<tr>
<td>11224</td>
</tr>
<tr>
<td>12213</td>
</tr>
<tr>
<td>12115</td>
</tr>
<tr>
<td>22115</td>
</tr>
<tr>
<td>21114</td>
</tr>
<tr>
<td>21124</td>
</tr>
<tr>
<td>11133</td>
</tr>
</tbody>
</table>

The score minus the negative overlap score gave the final overlap measure. The mean overlap score for the generalize transfer items was 23.6 and for the control transfer items was 18.6. Since this difference was somewhat larger than we had hoped to achieve, we will examine performance on specific items paired in such a way as to contrast the effect of higher similarity to study materials without generalizations with the effect of lower similarity to study materials with generalizations.

In generalization, control, or control, was varied within subjects. The experiment was run in two phases. The order of phases, generalize-control-control-generalize, and the assignment of club member or space invader descriptions to generalize and control materials were counterbalanced across subjects.

The procedure was identical to that of Experiment 2. Both generalize and control study items were presented in a blocked order, using the method described in Experiment 2. After each transfer test, subjects filled in brief questionnaires in which they described what strategies they used to learn the study items and their impressions of what determined category membership.

Results and Discussion

There were 26 cases of failure to reach learning criterion. Five subjects did not reach criterion for either their generalize or control phase. Of the remaining 16 cases, 4 occurred in the generalize phase and 12 occurred in the control phase. Subjects took 9.55 trials to reach learning criterion with generalize materials and 10.68 trials with control materials. This effect approached significance, F(1, 38) = 3.65, p = .06. Learning was faster in the second phase, regardless of materials, as revealed by a significant Study Material × Phase Order interaction, F(1, 38) = 13.01, p = .001. Generalize-control subjects took 10.3 trials in their first (generalize) phase and 9.3 trials in their second (control) phase. Control-generalize subjects took 12.1 trials in their first (control) phase and 8.8 trials in their second (generalize) phase. The speedup across phases for subjects going from generalize to control materials was one trial, but for subjects going from control to generalize materials the decrease in learning time was more than three trials. These learning phase data replicate the findings of Experiment 1. Learning was facilitated when the study items afforded category generalizations, even when the two sets of study materials had approximately equal interitem similarity.

For generalize materials, the mean accuracy and confidence scores on study items at retest was 82% and 2.87, respectively. For control materials, these scores were 82% and 2.91.

Table 7 presents the mean accuracy and confidence scores as a function of phase order and materials for the transfer items. There was a main effect of item type (study vs. transfer) on accuracy, F(1, 38) = 25.1, p < .001. Not surprisingly, subjects were less accurate on new transfer items than on studied items, indicating some effect of memory for specific instances. Although subjects were equally accurate on generalize and control study items, they differed significantly in accuracy on the three-overlap transfer items as a function of study materials (p < .05). With generalize materials, subjects were 78% accurate on transfer items, whereas with control materials, they were 71% accurate on transfer items. Phase order did not significantly affect accuracy, but the Phase Order × Study Materials interaction approached significance, F(1, 38) = 3.5, p < .06. This reflected the trend that control-generalize subjects were 12% more accurate in their second (generalize) transfer test than they were in their first (control) transfer test, whereas generalize-control subjects were 2% more accurate in their second (control) phase than in their first (generalize) phase. At the very least, these data suggest
that the benefit of practice with the task is contingent on the nature of the materials.

Similar effects emerged for confidence ratings. The item effect was significant, \( F(1, 38) = 30.3, p < .001 \). The mean confidence score on transfer items for generalization materials, 2.50, was significantly higher than the mean confidence score on these items given control materials, 1.85 \( (p < .05) \). The interaction of study materials with item type was significant, \( F(1, 38) = 8.8, p = .005 \). Study material interacted with phase order, \( F(1, 38) = 6.2, p = .017 \). The mean increase in confidence on the second phase relative to the first was 1.11 for control-generalize phase order, but only .16 for the generalize-control phase order.

As in Experiment 1, we examined performance on specific items. Table A3 in the Appendix gives accuracy and confidence means for individual transfer items collapsed across subjects and averaged over Clubs 1 and 2. We found 10 pairs (5 Club 1 pairs and 5 corresponding Club 2 pairs) such that the similarity of the control item was nearly identical to the similarity of the generalize items. The average similarity of the control pairs was 21.6 and of the generalize pairs was 21.2. The average accuracy for the control items in these pairs was 74% and for the generalize items was 78%. The average confidence rating for the control items was 2.00; for the generalize items it was 2.40. The accuracy effect was not significant, but the confidence effect was significant across items, \( r(9) = 1.86, p < .05 \), and across subjects, \( r(19) = 1.98, p < .05 \). Thus it appears again that when we compensate for effects of item similarity, we still find an advantage for generalizations.

When we examined the subjects’ postexperimental reports, we found little evidence for awareness of generalizations or parts of generalizations. There was certainly no case in which a subject reported all six three-feature generalizations that occurred in his or her study items. When asked what determined category membership, most subjects listed single features. A few subjects showed sensitivity to configurations of features or contingency relationships (e.g., “Dolphin members were Lutheran and collected stamps, unless they liked jazz... then they were Koalas”). We checked subjects’ reports for the generalize phase to see how well their rules matched the six generalizations that actually appeared in their study items. Out of the 30 subjects for which we had protocols, one subject reported two complete generalizations; another subject mentioned one. There were seven subjects reporting two thirds of some of the generalizations. However, these subjects, like the majority, also reported feature combinations that did not correspond at all to the generalizations. In general, subjects were either unaware of the category generalizations or unable to articulate them.

**General Discussion**

Perhaps the best testimony to the success of these experiments is that none of the theories we reviewed in the introduction has emerged unscathed. Experiments 1 and 3 provided ample evidence that the opportunity to form category generalizations leads to better performance in initial learning and transfer. While Experiment 2 did not directly contrast a control versus generalize condition, the contrast between blocked and random presentation would only have impact if subjects were forming generalizations.

---

*Percent correct.

**Table 7**

<table>
<thead>
<tr>
<th>Study condition</th>
<th>Phase order</th>
<th>Generalize-Control</th>
<th>Control-Generalize</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalize</td>
<td>Accuracy*</td>
<td>74</td>
<td>82</td>
<td>78</td>
</tr>
<tr>
<td>Control</td>
<td>72</td>
<td>70</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Generalize</td>
<td>Confidence</td>
<td>2.11</td>
<td>2.90</td>
<td>2.50</td>
</tr>
<tr>
<td>Control</td>
<td>1.95</td>
<td>1.79</td>
<td>1.85</td>
<td></td>
</tr>
</tbody>
</table>

*Since there were significant order effects, scores of subjects who had the control-first-generalize-second order were averaged with scores of subjects who had the generalize-first-control-second order. A given subject’s scores were averaged with the next subject run in the alternative order. We then did our subject analysis on these 20 paired subjects.*
prototype models suppose that central tendencies are extracted and used to categorize test instances. However, they assume a single central tendency that implies that distance from the central tendency should be the only relevant variable. Numerous experiments have already disconfirmed this prediction, at least for cases in which verbal material is used (e.g., Hayes-Roth & Hayes-Roth, 1977) or for cases in which small categories are learned (Medin & Schaffer, 1978). Experiment 1 showed that, for our particular paradigm too, there is an effect of degree of overlap with individual study instances, holding number of diagnostic features (i.e., central tendency) constant.

Our results also rule out most of the feature-set models (Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974; Reitman & Bower, 1973) in that they have no role for a generalization process. Both Neumann’s (1974) model and Reitman and Bower’s (1973) model were designed to predict recognition ratings; neither makes classification predictions. Hayes-Roth and Hayes-Roth’s (1977) property set model, which is most similar to the generalization model we tested, does not predict the differences we found between generalize and control conditions. Their model predicts classification on the basis of a transfer item’s most diagnostic property set. According to their model, an item’s property sets are the power set of all its values. Each of our five-feature items had 31 property sets. To make property set model predictions for our Experiment 3 task, we did the following for each property set of each transfer item: (a) tabulated its frequency of occurrence among Category 1 exemplars and among Category 2 exemplars; (b) formed two ratios—the frequency of Category 1 occurrences over all occurrences and the frequency of Category 2 occurrences over all occurrences; and (c) found the largest ratio of all those computed, which signified the most diagnostic property set. If it is a Category 1 ratio, the model predicts a Category 1 classification for the transfer item. If it is a Category 2 ratio, the model predicts a Category 2 classification. If there is a tie for the largest ratio across categories, classification is not predicted.

In Experiment 3, there are 12 generalize transfer items with a most diagnostic property set for the correct category. There are 10 such control items. Performance is 81% accurate on these generalize items and 73% on these control items. Thus, when we consider only the transfer items for which the property set model makes a classification prediction, it still cannot account for the difference found between generalize and control conditions.

This leaves only an instance-only model and the ACT generalization model, the two alternative views we set out to differentiate. As indicated by the various problems we encountered in constructing material to this end, this was no easy task. Let us first consider the body of evidence from all three experiments that supports the generalization model: (a) Generalizable materials were learned faster than nongeneralizable materials (Experiments 1 and 3), even when the nongeneralizable materials had a higher degree of interitem similarity (Experiment 1); (b) generalizable material was learned faster when presented so as to facilitate the formation of generalizations (Experiment 2); (c) transfer performance was better when generalizable materials were blocked to facilitate the formation of generalizable materials than when they were presented randomly (Experiment 2); (d) performance was higher on items classifiable by generalizations than on items that had higher similarity to the study material but were not classifiable by category generalizations (Experiments 1 and 2).

On the other hand, the data from the experiments gave ample evidence for effects of similarity between study and transfer items even when there were applicable generalizations. These results cannot be explained by the generalization model as set forth in Anderson et al. (1979) but can be explained by an instance-similarity model.

Is there any way in which an instance-based model can account for the above evidence in favor of a generalization model? One possibility lies in Medin and Schaffer’s (1978) model, which includes a selective attention mechanism. The purpose of this mechanism is to account for the differential saliency of certain dimensions for different subjects and the possibility that subjects may actively engage in some hypothesis testing. In either case, the final representational out-
GENRALIZATIONS AND SCHEMA ABSTRACTION

The major inadequacy with the ACT theory as formulated by Anderson et al. (1979) is its failure to allow items to be classified on the basis of partial matches to generalizations. In response to these results and other considerations, the pattern matching assumptions have been reformulated to permit partial matches to both specific instances and to category generalizations. This means that there are two ways in which similarity can aid classification of novel items. First, interitem similarity can lead to category generalizations at study. In addition, similarity between transfer items and either specific study items or higher order category information (generalizations) can serve as a direct basis for categorization. In fact, the current ACT model uses the same partial matching techniques for detecting similarities between study items to form generalizations as it uses for categorizing new items.

The reformulation of the model to allow partial matches is not just a concession to evidence that similarity determines transfer performance to a certain extent. A similarity detecting process can presumably operate on any stored pattern and we see no reason to limit it to only specific item representations. The original issue was, of course, whether any higher order feature patterns are even formed and, if so, whether they determine performance on tasks commonly used to investigate concept formation. We believe, given a category with a large number of exemplars, classification of novel items is in part determined by generalizations formed from study items. There are just too many pieces of data across our three experiments to doubt this. However, there is equally strong evidence indicating that item similarity also influences item classification.

Thus, the present data suggest that it may be both unnecessary and inappropriate for a theory of schema abstraction to choose...
between rule abstraction mechanisms and analogy (similarity to instances) mechanisms. They need not be mutually exclusive processes. Given the evidence that people do recall specific items and use them to classify novel items plus the evidence that specific items give rise to category generalizations that facilitate classification of novel items, it seems more interesting and perhaps fruitful to regard these processes as complementary rather than competing and to try to incorporate them into a single parsimonious model. Both specific instance information and higher order generalizations may be available for a partial-matching mechanism to operate in the same way on each type of information.

References

Appendix

Table A1
Similarity Scores, Mean Accuracy, and Mean Confidence for Experiment 1 Transfer Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Similarity score</th>
<th>Accuracy</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Generalize</td>
<td>Control</td>
<td>Generalize</td>
</tr>
<tr>
<td>Three-overlap</td>
<td>11313</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>11413</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>14322</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>14422</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>41111</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>41211</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>22142</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>22242</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Two-overlap</td>
<td>12413</td>
<td>15</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>12313</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>41422</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>41322</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>13111</td>
<td>7</td>
<td>-7</td>
</tr>
<tr>
<td></td>
<td>13221</td>
<td>7</td>
<td>-9</td>
</tr>
<tr>
<td></td>
<td>12141</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>12241</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>

* Averaged over Club 1 and Club 2 items, blocked and random conditions.
### Table A2
**Experiment 2: Set B Generalize Study Items and Transfer Items**

<table>
<thead>
<tr>
<th>Study items</th>
<th>Four-overlap</th>
<th></th>
<th></th>
<th>Three-overlap</th>
<th></th>
<th></th>
<th>Two-overlap*</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>Category 2</td>
<td>Category 1</td>
<td>Category 2</td>
<td>Category 1</td>
<td>Category 2</td>
<td>Category 1</td>
<td>Category 2</td>
<td>Category 1</td>
<td>Category 2</td>
</tr>
<tr>
<td>11321</td>
<td>44234</td>
<td>11111</td>
<td>44444</td>
<td>11151</td>
<td>44454</td>
<td>*13112</td>
<td>*42442</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11441</td>
<td>44114</td>
<td>11141</td>
<td>44444</td>
<td>11161</td>
<td>44464</td>
<td>*21112</td>
<td>*34443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11211</td>
<td>44334</td>
<td>11341</td>
<td>44214</td>
<td>42125</td>
<td>13435</td>
<td>*21112</td>
<td>*34443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43121</td>
<td>12431</td>
<td>42121</td>
<td>13434</td>
<td>46122</td>
<td>16433</td>
<td>*24212</td>
<td>*31343</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44122</td>
<td>11433</td>
<td>41122</td>
<td>14433</td>
<td>52231</td>
<td>53322</td>
<td>23442</td>
<td>33113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41124</td>
<td>14431</td>
<td>42122</td>
<td>13433</td>
<td>62231</td>
<td>63324</td>
<td>21143</td>
<td>34412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22233</td>
<td>33322</td>
<td>22232</td>
<td>33323</td>
<td>23132</td>
<td>32423</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Asterisks indicate items which partially match a category generalization.

### Table A3
**Similarity Scores, Mean Accuracy, and Mean Confidence for Experiment 3 Transfer Items**

<table>
<thead>
<tr>
<th>Item</th>
<th>Similarity score</th>
<th>Accuracy</th>
<th>Confidence</th>
<th>Similarity score</th>
<th>Accuracy</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Generalization condition</td>
<td></td>
<td></td>
<td>Control condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11224</td>
<td>17</td>
<td>.83</td>
<td>3.21</td>
<td>11241</td>
<td>18</td>
<td>.68</td>
</tr>
<tr>
<td>11213</td>
<td>28</td>
<td>.79</td>
<td>2.62</td>
<td>12135</td>
<td>22</td>
<td>.77</td>
</tr>
<tr>
<td>12213</td>
<td>26</td>
<td>.80</td>
<td>2.61</td>
<td>21123</td>
<td>25</td>
<td>.77</td>
</tr>
<tr>
<td>12115</td>
<td>27</td>
<td>.79</td>
<td>2.34</td>
<td>11314</td>
<td>19</td>
<td>.71</td>
</tr>
<tr>
<td>22115</td>
<td>23</td>
<td>.79</td>
<td>2.32</td>
<td>41111</td>
<td>24</td>
<td>.77</td>
</tr>
<tr>
<td>21114</td>
<td>21</td>
<td>.72</td>
<td>2.09</td>
<td>32122</td>
<td>18</td>
<td>.71</td>
</tr>
<tr>
<td>21124</td>
<td>19</td>
<td>.74</td>
<td>2.27</td>
<td>23211</td>
<td>18</td>
<td>.76</td>
</tr>
<tr>
<td>11125</td>
<td>28</td>
<td>.83</td>
<td>2.74</td>
<td>14412</td>
<td>5</td>
<td>.58</td>
</tr>
</tbody>
</table>

Received September 15, 1980
Revision Received April 21, 1981