

The effects of information order and learning mode on schema abstraction

RENÉE ELIO and JOHN R. ANDERSON
Carnegie-Mellon University, Pittsburgh, Pennsylvania

Three experiments investigated the effects of information order and representativeness on schema abstraction in a category learning task. A set of category members, in which the variability and frequency of member types were correlated, was divided into four study samples. In the high-variance condition, each sample was representative of the allowable variation in the category and the frequency with which it occurred. In the low-variance condition, the initial study sample focused only on the most frequently occurring category members. Subsequent samples gradually introduced exemplars, and hence additional variance, from remaining member types. After the fourth study sample, all subjects in all conditions had seen the same category members. Experiment 1 revealed that transfer performance was better if subjects began with a low-variance sample and were gradually introduced to the allowable variation on subsequent samples than if they consistently saw representative samples. Experiments 2 and 3 suggested that this information-order effect may interact with learning mode: Subjects induced to be more analytic about the material performed better if their initial and subsequent samples were representative of the category variation.

To a large extent, learning involves the incorporation of new information into some existing knowledge structure. A learner's first exposure to some domain may determine the nature of that structure, which in turn can influence how subsequent information in that domain is processed and incorporated into what is already known. This study investigated how abstraction of information about ill-defined categories is affected by varying the nature of initial category exemplars that a learner encounters and how subsequent exemplars are introduced.

The idea that the order in which information is received could affect both the learning process and the ultimate knowledge representation is not particularly new. General learning theories, such as Rumelhart and Norman's (1978) model of accretion, tuning, and restructuring, as well as social cognition models of impression formation (N. R. Anderson, 1968; Asch, 1946), are sensitive to the notion that initial information can affect the manner in which the learner incorporates subsequent knowledge into what he or she already knows. Early concept-identification research demonstrated that information order can affect the discovery of simple classification rules. Bruner, Goodnow, and Austin (1956) suggested that the learner uses some

aspects of the first instances encountered to form a set of hypotheses. Subsequent information then serves as confirming or infirming evidence for what the learner initially hypothesized as being potentially relevant. Other concept-identification studies have also shown that the order in which category exemplars and non-exemplars are presented affects how hypotheses are tested, retained, or rejected (Hovland & Weiss, 1953).

These concept-identification studies clearly demonstrated the importance of information order on decision-making and hypothesis-testing processes. However, many of the categories people encounter do not follow the kind of easily specified and tested rules that characterize most of the early concept-identification research. More recently, attention has been given to how people learn ill-defined categories—categories for which there is no single, easily specified rule that is singularly predictive of category membership. Since information about such categories does not reduce to a single rule, investigating the nature of that information and how it is represented has been the focus of considerable experimental effort. We will refer to the process by which such categories are learned from experience with category exemplars as schema abstraction. Schema-abstraction experiments usually employ a two-stage paradigm consisting of a study or training phase followed by a transfer phase. In the study phase, a subject either learns to classify a set of training items to criterion or simply has some fixed exposure to them and their category membership. In the transfer phase, subjects are asked to categorize or recognize items that include novel instances that they have not studied. It is their transfer performance on the novel items that is used to

This research was supported in part by an NSF graduate fellowship to the first author, who is now at the Alberta Research Council, and ONR Contract N00014-81-0335 and NSF Grant IST-80-15357 to the second author. We would like to thank two anonymous reviewers for their helpful comments and suggestions. Reprint requests should be sent to Renée Elio, Computing Department, Alberta Research Council, 11315 87th Avenue, Edmonton, Alberta T6G 2C2, Canada.

evaluate the nature of the category information acquired during experience with the training items.

Most schema-abstraction experiments have manipulated the nature of the training items in order to differentiate alternative models. There has been little investigation of factors affecting the acquisition of the original training items. In general, whether a schema-abstraction theory assumes that classification of novel items is based on similarity to stored instances (Medin & Schaffer, 1978), a central tendency or prototype (Franks & Bransford, 1971; Posner & Keele, 1968), or the frequency of feature and feature combinations during learning (Elio & Anderson, 1981; Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974; Reitman & Bower, 1973), a default prediction seems to be that every study item encountered will have an equal impact on the final schema and this impact will be independent of exemplar presentation order.

The nature of the exemplar training set as a whole (e.g., its size and the variability of included items) has been manipulated in several studies (Barresi, Robbins, & Shain, 1975; Homa, 1978; Homa & Vosburgh, 1976; Posner & Keele, 1968). Posner and Keele presented subjects with either a high-variability or a low-variability exemplar training set and predicted that experience with a low-variance set would facilitate abstraction of the central tendency. Although low-variance material was learned faster, training experience with high-variance exemplars led to the best transfer. Using visual stimuli, Homa and Vosburgh demonstrated that, if the to-be-learned category is large and variable, then transfer performance is facilitated by giving the learner a large exemplar training set that reflects the amount of allowable variation in the category. Along similar lines, Fried and Holyoak (Note 1) proposed that a learner induces a representation of the distribution of category members across a feature space, based on the sample given for training. According to their model, a nonrepresentative sample not only hinders learning, but also reduces the probability that an initial misconception can be corrected. This concurs with Homa and Vosburgh's finding that a sample representative of category variance leads to better transfer than one that does not reflect the variation that is encountered during a transfer test.

It seems, however, that most schema-abstraction theories, whether they fall into the class of prototype, strength, or exemplar models, would predict better transfer performance given a training set that reflected all the critical characteristics of the category and the range of variation on specific dimensions. In other words, it seems unreasonable to expect particularly good classification of penguins and ostriches as birds given a training set of robins, blue jays, and sparrows. Those studies that have manipulated the nature of the training material have demonstrated that schema abstraction is influenced by "what" is encountered, for example, high-variance, representative material versus low-variance, nonrepresentative material. In the set of experi-

ments reported here, all subjects received the same set of training material and, hence, encountered the same degree of category variance. The manipulations involved when and how they encountered that variance.

GENERAL METHOD

Our general goal was to evaluate the effects of information order and variance on schema abstraction by measuring transfer performance after different numbers and different types of category exemplars had been studied. By the end of Experiment 1, all our subjects had been trained on the same category members, and, as a consequence, the subjects in all conditions encountered the same amount of category variability. What we manipulated was how the set of category members was divided into study samples, how these were ordered for learning, and, therefore, how category variance was encountered.

Our stimuli were descriptions of people belonging to one of two clubs. A club member's description was created from five 4-valued dimensions. Each item could be symbolically represented as five digits, one for each characteristic of the person. The value of each digit represented the value on that dimension. For example, an item such as 12113 might translate as "works for the government, is college educated, is single, plays chess, likes jazz." We constructed our categories using this numerical item notation. Its correspondence to verbal descriptions for our subjects will be explained in greater detail below.

In order to test how schema abstraction might be affected by how the category variance was encountered, a large, variable category in which different member types both shared many overlapping features and had unique feature patterns was needed. We could then manipulate exposure to category variance by exposing subjects to all member types equally frequently or by concentrating on a subset of these types.

One category's items was arbitrarily defined to have a majority of dimensions with values 1 and 2; the alternative category's items had a majority of dimensions with values 3 and 4. To generate the first category, we defined prototype feature patterns that differed in which and how many dimensions had value 1 or value 2. Type A prototype feature patterns had a single dimension in any position with value 2 (21111, 12111, 11211, 11121, and 11112). Type B prototype feature patterns had two dimensions with value 2 in positions 1-3 (22111 and 12211). Type B' prototype feature patterns had two dimensions with value 2 in position 3-5 (11221 and 11122). Type C had three dimensions with value 2 in positions 1-3 (22211); Type C' had three dimensions with value 2 in positions 3-5 (11222).

From these prototype feature patterns, the actual category members were generated by changing a dimension with value 1 to have either value 3 or value 4. The set of all possible category members that could be gen-

erated from these particular prototype patterns would be too large for subjects to learn. Therefore, only a subset of possible members was used for each type, with the constraint that values 3 and 4 each appear approximately the same number of times for each dimension. The members for category 1 are given in Table 1. The members for category 2 were created by interchanging values 1 and 4 and interchanging values 2 and 3. Thus, for the category-1 item 12213, there was a corresponding category-2 item 43342.

Even with the constraints that different member types have one, two, or three dimensions with value 2, there are many more possible prototype patterns that we did not use. We selected this particular set of prototype feature patterns to create a framework of overlapping feature patterns between different member types. For example, although Type C items do not have many overlapping features with the majority of other items in the category, they share similar feature patterns with Type B items.

The general aim was to give all subjects the same exposure to category variance while manipulating the order in which they encountered that variance. To accomplish this, the set of category members was divided into four equal study samples, all of which would be presented to a subject during the course of the experiment. These samples differed in which member types were included and how many from each type were included. Different sampling conditions were defined in terms of the initial sample's characteristics, since this was a subject's first exposure to the category. There were two primary samp-

ling conditions. In the high-variance condition, the initial sample contained items from all five member types in frequencies proportional to their occurrence in the whole category. In this respect, it was representative of both the type of variation in the category and the frequency with which that variation occurred. In the other sampling condition, the initial sample was a low-variance introduction to the category, because most of its items were drawn from the member types with the most members (e.g., Type A). If this were the subject's only experience with the category, the method would be nearly identical to that in previous studies that manipulated the variance of a single training set. However, subsequent samples gradually introduced the remaining category members. We measured transfer performance after each sample had been studied, but the main interest was final performance following the fourth sample, at which point all subjects in all conditions had studied all the category members. On the basis of previous studies (Homa & Vosburgh, 1976; Posner & Keele, 1968) and of our own intuitions, we predicted that subjects given representative samples of category variability would do less well initially but ultimately would perform best.

EXPERIMENT 1

Method

Stimuli. The category items were descriptions of people who belonged to one of two clubs. These descriptions were generated from five 4-valued dimensions: job—(1) unemployed, (2) self-employed, (3) government, (4) private firm; marital status—(1) single, (2) married, (3) divorced, (4) widowed; education—(1) grammar school, (2) high school, (3) college, (4) trade school; hobby—(1) tennis, (2) chess, (3) golf, (4) bowling; musical taste—(1) classical, (2) jazz, (3) rock, (4) bluegrass. Each item could be symbolically described as five digits, one for each dimension. The value of the digit represented the value on that dimension. For example, using the order of dimensions and the order of values within dimensions given above, the item 23411 would correspond to the description "self-employed, divorced, trade school, tennis, classical." Both the order of dimensions and the order of values within a dimension were determined randomly for each subject, so that this same item 23411 might be instantiated as "jazz, college, golf, single, government" for another subject.

Design. Each category was divided into four samples, according to the following selection rules. Each rule describes how many items from each of the five member types (C, B, A, B', and C') were included in a given sample. The frequency of items in each member type for the whole category was 4-8-16-8-4 for Member Types C, B, A, B', and C', respectively. By varying the number of items from each member type included in each sample, we could construct samples that differed in category variance and representativeness.

In the representative condition, each of the four samples contained items from Types C, B, A, B', and C' in frequencies proportional to their occurrence in the whole category: 1-2-4-2-1. Since all member types were included, these samples correspond to high-variance training sets that reflect all the category variation. In the centered condition, the initial sample concentrated on the member types with the most members, consisting of 0-1-8-1-0 items from Member Types C, B, A, B', and C'. This sample corresponds to a low-variance training set that does not

Table 1
Stimulus Items for Experiments 1-3

Member Type				
C	B	A	B'	C'
Prototypes				
22211	22111 12211	21111 12111 11211 11121 11112	11122 11221	11222
Category Members				
22231	22113	21113	31221	13222
22241	22114	21114	41221	14222
22213	22311	21131	13221	31222
22214	22411	21141	14221	41222
	12231	12311	11322	
	12241	12411	11422	
	12213	13211	31122	
	12214	11241	41122	
		31211		
		11214		
		11321		
		11421		
		31112		
		41112		
		13112		
		14112		

illustrate the allowable range of variation in the category. The second and third centered samples had frequencies of 1-3-2-3-1 and 2-2-2-2-2 of Member Types C, B, A, B', and C'. These samples served to introduce member types and, hence, category variation excluded from the initial sample. A variant of the centered condition, the late-centered condition, reversed the centered condition samples 1 and 3, so that the concentration on Type A occurred later during learning, rather than initially. All conditions used a representative sample as the fourth and last sample in the series. The items in this representative sample were fixed across subjects and across conditions.

The 10 items in each sample were selected randomly from Club 1 according to the sampling constraints. The corresponding 10 items from Club 2 were then included, so that each sample of items contained 20 items altogether. By the end of the fourth block, the four samples had introduced all 40 members of each club.

Subjects. Forty-five male and female students from the Carnegie-Mellon University community participated as subjects; they received credit toward an optional psychology course assignment and/or \$3.00 per hour. There were 15 subjects in each of three experimental conditions. The experimental session lasted 2 h.

Apparatus and Procedure. The experiment was controlled by a PDP-11/34 computer. The subjects were run in groups of one to five. Each subject sat in a room that contained a CRT screen on which items were presented. The subjects entered their responses using the terminal keyboard.

The experiment was divided into four study-test blocks. These four blocks corresponded to the four study samples designed for each condition. In the study phase of each block, the 20 items of that block's sample were presented 1 at a time on the screen. Each study phase consisted of three passes through the 20 items, although the subjects were told that they would keep cycling through the 20 items until they had reached an undisclosed criterion. The subjects were instructed to classify each item as either a "Dolphin Club" or a "Koala Club" member (these names were chosen to correspond to the keyboard keys "d" and "k," used as the response keys). They received immediate feedback, of the form "Right/Wrong, Dolphin/Koala Club." The feedback and the description remained on the screen for 10 sec. The screen was then erased, and the next item was presented. If the subjects did not respond within 10 sec of item onset, the item's club membership appeared automatically and remained on the screen with the item for 10 sec. The subjects were informed that failure to respond within 10 sec counted as an error. After each pass through the 20 items, the subjects were told their accuracy.

After the third study pass through the block's study sample, the subjects moved to a transfer test.¹ The transfer-item set consisted of the current block's 20 study items, the 20 never studied prototypes, and 40 items (20 from Club 1 and the corresponding 20 from Club 2) drawn randomly from remaining club members. Transfer items appeared on the screen 1 at a time, and the subjects were instructed to classify each item as quickly as possible. The subjects did not receive feedback on their classifications. After they had classified an item, they entered a typicality rating from 1 to 5, where 1 meant "not very typical of the club [to which the subject had assigned it]" and 5 meant "very typical of the club [to which the subject had assigned it]." The screen was erased immediately after the typicality rating had been entered, and the next item was presented.

The experiment proceeded as an alternation between study phases, during which a new sample was presented for learning, and transfer phases. For each transfer test, study items from previous blocks' study samples were reinserted into the item pool from which the transfer items for the current test could be drawn. By the end of the experiment, all the items in both categories had appeared once in a study phase (with feedback).

Results

Mean accuracy and typicality ratings were computed for each member type for each subject. Typicality scores for a given member type were computed as the mean of a subject's typicality ratings on correct classifications minus his or her typicality ratings on incorrect classifications of items within that member type. These typicality scores ranged from -5 to +5.

Table 2 presents the mean accuracy and typicality for the last (block 4) transfer test. The effect of sampling conditions was significant [$F(2,42) = 4.97, p < .02$]. Newman-Keuls tests indicated that centered condition subjects gave significantly higher typicality ratings than did subjects in the other two conditions.² The accuracy means presented in Table 2 followed the same pattern as the typicality ratings, although the effect only approached significance [$F(2,42) = 2.54, p = .09$].

In addition to the category members, each transfer test included the prototypes used to generate category members, although these prototypes were never included in any study phase. The lower portion of Table 2 gives the prototype accuracy and typicality means as a function of sampling condition for the last transfer test. There was a significant sampling effect on typicality ratings given to the prototypes [$F(2,42) = 4.49, p < .02$]. Newman-Keuls tests indicated that the mean typicality rating given in the representative condition was significantly lower than those given in all other conditions. The centered-condition typicality mean was significantly larger than the late-centered-condition mean. There were no main effects of sampling on accuracy [$F(2,42) = 1.65, p = .20$] and no main effects of member type.

Overall, the final transfer performance indicates an advantage for beginning with a low-variance sample, such as that in the centered condition, rather than a high-variance sample that is representative of the allowable category variation. However, the block-by-block transfer performance on specific member types provides insight into how these final performance differences evolved. Table 3 gives the mean typicality scores on studied items for the three conditions as a function of block and member type. For simplicity, the less frequent

Table 2
Mean Block-4 Accuracy and Typicality Ratings as a Function of Condition in Experiment 1

	Sampling Condition			Mean
	Representative	Centered	Late-Centered	
	Studied Items			
Typicality	2.22	3.55	2.70	2.82
Accuracy	.80	.91	.84	.85
	Prototypes			
Typicality	2.68	4.25	3.59	3.51
Accuracy	.85	.96	.93	.91

Member Types B, B', C, and C' were collapsed into one group for presentation in Table 3. There were significant main effects of block and member type and significant block x member type interactions ($p_s < .001$) in all cases. Representative subjects showed an uneven but moderate improvement over blocks. In contrast, accuracy in the centered condition rose consistently on every block and every member type. The block-1 accuracy and typicality means for the centered condition showed some effect of the initial focus on Type A items. With the introduction of Member Types B-C and B'-C' on blocks 2 and 3, there was a corresponding rise in accuracy and typicality ratings for these items. Note that the originally high accuracy on Type A in the centered condition remained unaffected by the gradual exposure to other member types. In fact, it continued to rise on blocks 2 and 3, although the study samples for these blocks contained very few Type A items.

Is the relatively poor showing of the representative condition due solely to an initial "bad start"? Examining the degree of improvement over blocks is one way to evaluate this possibility. First, there was no advantage of representative over centered presentation in the degree of improvement on any specific member type. When performance is collapsed across member types, subjects in both the centered and the late-centered conditions improved their accuracy by 13%, whereas subjects in the representative condition improved by 7%. Second, the poor performance of subjects in the late-centered condition relative to that of subjects in the centered condition indicates that the beneficial effects found with a low-variance sample depend on encountering that sample initially, rather than at any time, during learning.

Since all the items eventually appeared in study phases, the distinction between transfer items and study items differs from what characterizes the usual schema-abstraction paradigm. That is, inferences based on transfer performance usually depend on subjects' reaching equivalent degrees of learning on training items. However, the differential acquisition of study information is part of the phenomenon we are investigating. It was possible to separate performance on the last transfer test into that block's immediately preceding study items and its current transfer items. Under this breakdown of data, the conditions still differed on

typicality ratings given to items that had not just appeared in the preceding study sample [$F(2,42) = 4.16$, $p = .02$]. The mean typicality ratings for these items were 2.20, 2.56, and 3.45 in the representative, late-centered, and centered conditions, respectively. Centered-condition subjects were also 10% more accurate than the representative-condition subjects on these items (89% vs. 79%, respectively), although the overall main effect of sampling on accuracy was not significant.

Discussion

The overall finding of Experiment 1 was that the abstracting of category information is facilitated by beginning with a low-variance sample rather than with one that reflects the amount of allowable variation that will be encountered. The fact that accuracy effects are only marginally significant is not disturbing in light of the substantial typicality-rating differences. Rosch and Mervis (1975) found typicality ratings to be a sensitive measure of how natural categories are represented. The two categories we used were fairly discriminable; accuracy performance was high even in the worst cases. We believe that the significant differences found in typicality ratings, with consistent supporting trends in accuracy, indicate that the final conceptions of these categories were different for different sampling conditions.

As noted earlier, the transfer tasks were unique in that all so-called transfer items had been studied with feedback once during the course of the experiment. It is important to consider whether the subjects' performance reflects differences in item memory, rather than abstraction: Did centered-condition subjects just learn a larger number of items to a better degree than did representative-condition subjects? A number of findings suggest this was not the case. If the paradigm evaluated only differences in item memory, then differences in prototype classification would not be expected. Second, if the high degree of interitem similarity within low-variance training samples facilitated item learning rather than abstraction, then it would be difficult to explain how reordering of samples (centered vs. late-centered) resulted in different performance. The contrast between the centered and late-centered conditions, which differed only in the presentation order of otherwise identical study samples, testifies to the impact of when the variance is encountered. For ex-

Table 3
Mean Typicality Ratings as a Function of Member Type, Block, and Sampling Condition for Experiment 1

Condition	Noncentered Items*				Centered Items**			
	Block				Block			
	1	2	3	4	1	2	3	4
Representative	1.82	2.05	1.90	2.36	1.55	1.64	1.84	1.96
Centered	1.97	3.04	3.61	3.71	2.75	2.81	3.37	3.24
Late-Centered	1.39	2.00	2.06	2.72	1.26	1.96	2.00	2.67

*Types B, B', C, and C'. **Type A.

ample, subjects in the late-centered condition also received a chance to learn the low-variance, frequently occurring Type A items in the block-3 sample. If the subjects were learning items rather than abstracting category information, then one might expect subjects in the late-centered condition to do better than centered-condition subjects on Type A items, since they encounter them more recently (on block 3 rather than on block 1). This did not occur.

The superiority of the centered condition over the representative condition suggests that an initial, low-variance sample of the most frequently occurring members may allow the learner to get a "fix" on what will account for most of the category members. If successful transfer hinges on exposure to the total amount of category variability in the training set, then we should have found a superiority of representative sampling in the transfer test after block 1.

There are several differences between this experiment and Homa and Vosburgh's (1976) study that may account for the differences in results. The first is their use of visual stimuli and our use of verbal stimuli. The extent to which any schema-abstraction model depends on a domain in which dimensions are continuous (usually visual) or discrete (usually verbal) is troublesome for interexperiment comparisons. Second, our centered-condition subjects did not see one type of category information (e.g., Type A items) to the exclusion of some other category information (e.g., Type C or Type C' items). In contrast, Homa and Vosburgh did exclude certain types of instances from their low-variance sample. This is a more easily corrected difference between our study and theirs. Specifically, since the items in the centered-condition sample did demonstrate that any value could appear on any dimension, it may not have been as "low variance," relative to Experiment 1's category, as Homa and Vosburgh's low-variance materials were, relative to their whole category. In an attempt to replicate their main finding with our materials, we ran a second experiment that included a fourth sampling condition. In this condition, the initial sample intensified the bias toward the most frequently occurring feature patterns by including only prototype patterns like those listed for Member Types B, A, and B' in Table 1. None of the items had dimensions with values 3 or 4 on them, and the ratio of value-1 dimensions to value-2 dimensions was either 4:1 or 3:2. Since so little of the category variability was evident in this sample, we expected transfer to be very poor given this initial training set.

Experiment 2 included this fourth sampling condition along with the other three described in Experiment 1. Since the block-by-block data from Experiment 1 suggested evolving category schemas, we asked our subjects directly what they thought governed category membership. During the course of Experiment 2, we obtained brief written protocols to provide additional insight into

how schemas were evolving differently for the various sampling conditions.

EXPERIMENT 2

Method

The design, stimuli, and procedure for Experiment 2 were exactly as described for Experiment 1, with the following exceptions. A fourth sampling condition, called the extreme condition, was included. The block-1 sample for this condition contained 10 randomly selected prototypes constructed similarly to those listed for Types A, B, and B' in Table 1. This set was increased with the items 21121, 21112, 12121, and 11212 to obtain a large enough item pool from which to draw the block-1 sample. Note that this group of subjects did not encounter all category members in a study phase by virtue of having studied these prototype-like items in block 1. Blocks 2-4 in the extreme condition were identical to the samples used for the centered condition.

The second change involved collecting written protocols from all subjects immediately after each transfer test and before the study phase for the next block. The subjects were given a questionnaire that asked them, "What is your current conception of the two clubs? What determines membership in the Dolphin Club? in the Koala Club?" Forty students, 10 in each of the four sampling conditions, participated in the study.

Results

Table 4 gives the overall accuracy and typicality means as a function of sampling condition for the final transfer test. Analyses of variance across all blocks revealed no significant differences among groups on typicality [$F(4,45) = 1.16, p = .34$] or accuracy [$F(4,45) = 1.46, p = .22$]. Additional analyses on block-4 performance, presented in Table 4, also revealed no significant effects of sampling. As we expected, however, the extreme condition did fare the worst, having the lowest mean typicality and accuracy ratings.

The mean accuracy and typicality ratings on prototypes as a function of condition are also given in Table 4. None of the differences among conditions on the final transfer test on any measure was statistically reliable.

The means in Table 4 indicate that the relative ordering of groups on each measure was markedly different from that found in Experiment 1. Table 5 provides a comparison of Experiment 1's and Experiment 2's

Table 4
Mean Block-4 Accuracy and Typicality Ratings as a Function of Condition in Experiment 2

	Sampling Condition				Mean
	Representative	Centered	Late-Centered	Extreme	
	Studied Items				
Typicality	2.80	2.74	2.58	2.14	2.57
Accuracy	.84	.86	.83	.76	.82
	Prototypes				
Typicality	3.80	3.81	3.82	3.80	3.81
Accuracy	.96	.96	.95	.95	.96

Table 5
Mean Block-4 Typicality and Accuracy Ratings for
Representative and Centered Conditions
in Experiments 1 and 2

	Sampling Condition	
	Representative	Centered
	Typicality	
Experiment 1	2.22	3.55
Experiment 2	2.80	2.74
	Accuracy	
Experiment 1	.80	.91
Experiment 2	.84	.86

results for the centered and representative conditions. Note that both accuracy and typicality means dropped for the centered conditions in Experiment 2 relative to those for the conditions in Experiment 1, whereas the representative-condition subjects gave somewhat higher typicality ratings in Experiment 2.

The protocols revealed little information about how subjects generally approach a task such as this and less about the impact of our sampling conditions. Only a few, general characteristics emerged. Almost all subjects described very gestalt-like views of the clubs (e.g., "Koalas seem to be a civil rights group"; "Dolphins are probably a singles' jazz appreciation club"). The extreme-condition subjects differed from the other subjects by (1) including the only reports (four cases) of a cue-validity strategy (e.g., "I counted up how many features there were for one club or the other and used that to classify a person") and (2) reporting changes in category rules and/or increasing difficulty as the experiment progressed (e.g., "Koala Club is letting in a higher class of people now [after block 2]"; "The reasons for classification seemed much more complex as the experiment continued [final comment after block 4]").

Discussion

The results of Experiment 2 were surprising, since it was identical in all respects to Experiment 1, save for the collection of protocols. However, the fact that we did not find the same results, but almost a reverse pattern, prompted us to reconsider the impact of protocol giving. Reber (1967, 1976) and Brooks (1978) have made distinctions between explicit and implicit learning and between analytic and nonanalytic modes of learning, respectively. "Analytic" mode has come to denote the conscious effort on the part of the subject to generate hypotheses and extract the regularities in a set of data. When a learner is in "nonanalytic" mode, he or she is presumed to be attending to the exemplars per se, encoding them rather than generating hypotheses about them. It is possible that the protocol-giving exercise may have induced our subjects to approach the task in a more analytic manner, which both Reber and Brooks have demonstrated influences performance in

tasks such as this. When we considered Experiments 1 and 2 as parts of a larger design, the possibility of an interaction between information order and learning mode was suggested: Namely, it may be better to begin with a low-variance sample if the training material was learned nonanalytically (Experiment 1), but with a high-variance, representative sample if the training material was learned nonanalytically. We designed a third experiment to test this interpretation and examined the effects of processing mode on only the centered and representative conditions, since these two sampling conditions offered the most theoretically interesting contrast. In Experiment 3, we crossed the sampling manipulation (representative or centered) with protocol giving (present or absent). Our interpretation of the first two experiments predicted an interaction between these two factors.

EXPERIMENT 3

Method

The sampling manipulation included only the representative and centered conditions described for Experiment 1. This factor was crossed with the rule manipulations: The subjects were either asked to indicate which features were relevant to club membership (rule condition) or were not (no-rule condition). This yielded four experimental conditions: representative-rule, representative-no rule, centered-rule, centered-no rule.

The procedure in the no-rule conditions was identical to that described for Experiment 1. The procedure for the rule condition was slightly different from that used in Experiment 2. Instead of the written protocol procedure used in Experiment 2, we presented each value of each dimension on the terminal screen and asked subjects to indicate whether they currently considered it relevant to determining membership in (1) the Dolphin Club, (2) the Koala Club, or (3) neither club. The subjects entered their responses by hitting the terminal keyboard keys "d," "k," or "n," respectively. Sixty-four students participated in the experiment, with 16 subjects in each of the four experimental conditions.

Results

Table 6 presents the mean accuracy and typicality on the final transfer test as a function of study condition and rule condition. Although the effects are small, the interaction of sampling and rule manipulations was significant for both accuracy [$F(1,60) = 4.63, p < .04$] and typicality [$F(1,60) = 6.93, p < .02$]. Performance was better in the rule condition given representative materials and better in the no-rule condition given centered materials. Both the accuracy and typicality interactions were significant for a transfer-item set that excluded the immediately preceding study items [$F(1,60) = 6.50$ and 8.15 , respectively, $ps < .02$]. There were no effects of either sampling or rule giving on prototype performance and no interactions.

Table 7 presents the typicality and accuracy data as a function of member type and block for the four experimental conditions. The most important observation to make from these data is the different impact the rule-giving mode had on typicality ratings over time

Table 6
Mean Block-4 Accuracy and Typicality Ratings as a Function of Rule and Sampling Condition in Experiment 3

Sampling Condition	Studied Items						Prototypes					
	Typicality			Accuracy			Typicality			Accuracy		
	R	NR	D	R	NR	D	R	NR	D	R	NR	D
Representative	2.82	2.17	+ .65	0.87	0.80	+ .07	2.46	2.59	-.13	0.94	0.94	0
Centered	2.13	2.64	-.51	0.77	0.83	-.06	2.41	2.41	0	0.94	0.94	0

Note—R = rule; NR = no rule; D = difference.

and member type for the two sampling conditions. In the representative condition, subjects in the rule condition maintained a constant advantage over subjects in the no-rule condition across member types and blocks. An analysis of variance on data from just these two conditions indicated that this advantage was significant [$F(1,30) = 4.74, p < .04$] and did not interact with block [$F(3,90) < 1$], member type [$F(5,150) = 1.11, p = .3$], or block and member type [$F(15,450) < 1$]. For the centered condition, the data in Table 7 indicated that the typicality ratings of subjects in the centered-rule condition dropped on Type A items (which had served as their initial introduction to the category) and rose slightly on the less frequently occurring items. The interaction of rule condition and member type was significant for the typicality data of subjects in the centered

condition [$F(15,450) = 1.79, p < .04$]. Although the apparent change in member-type typicality ratings over time for the two rule conditions was not statistically reliable, the trends in the Table 7 means suggest that subjects in the centered-rule condition initially gave about the same typicality ratings as did subjects in the centered/no-rule condition, but seemed to learn about the nature of the category more slowly than did their counterparts in the no-rule condition.

Rule-condition-subjects' block-by-block classifications of features as relevant or irrelevant determinants of club membership were scored for accuracy by comparing the relevancy judgment of a particular feature with that feature's association with one club or another. For each dimension, two of the values (e.g., those randomly assigned to the numbers 1 and 2 for a subject's particular stimulus set) were more often associated with the Dolphin Club and the other two values were more often associated with the Koala Club. We tabulated the frequency with which a subject correctly or incorrectly judged a given value of a dimension as relevant for determining membership in the club with which it was associated. The highest attainable score, collapsed across dimensions and blocks, was 80 (4 values x 5 dimensions x 4 blocks). The mean correct relevancy judgment was 36.2 for subjects in the representative-rule condition and 37.6 for subjects in the centered-rule condition. The mean incorrect judgement was 9.0 for subjects in the representative-rule condition and 13.3 for subjects in the centered-rule condition. Both groups of subjects assigned approximately the same number of features to the "relevant to neither club" category (27.3 and 29.1 for representative-rule and centered-rule, respectively).

Table 7
Mean Accuracy and Typicality as a Function of Member Type and Block for Experiment 3

	Member Type			
	No-Rule Condition		Rule Condition	
	NC	C	NC	C
Typicality				
Representative				
Block 1	1.53	1.91	1.95	2.18
Block 2	1.97	2.01	2.62	2.51
Block 3	1.90	2.16	2.60	2.63
Block 4	2.23	2.05	2.97	2.55
Centered				
Block 1	1.56	2.08	1.63	2.13
Block 2	2.19	2.32	1.74	1.78
Block 3	2.58	2.39	1.90	1.65
Block 4	2.57	2.79	2.24	1.92
Accuracy				
Representative				
Block 1	.73	.77	.75	.79
Block 2	.77	.79	.86	.84
Block 3	.76	.81	.85	.84
Block 4	.82	.80	.90	.83
Centered				
Block 1	.71	.76	.69	.80
Block 2	.80	.81	.74	.74
Block 3	.82	.82	.76	.74
Block 4	.84	.85	.80	.76

Note—NC = noncentered Types B, B', C, and C'; C = centered Type A.

Discussion

Experiment 3 provided some confirming evidence for an interaction of information representativeness and learning mode. These results suggest that, if a learner is in analytic mode, it may be best to present samples that are representative of the allowable variation. Conversely, if the learner is approaching the task nonanalytically, it may be best to present a low-variance sample and gradually to introduce variation. The small size of the interaction for category members and the absence of an interaction for prototypes is troublesome: The overall level of performance of subjects in the centered/

no-rule condition is notably lower than that found in Experiment 1. This may be due to a subject-sampling effect. Also, the method of protocol collection in Experiment 3 may have lessened the degree to which subjects were analytically considering the category exemplars by emphasizing individual features, rather than feature combinations.

GENERAL DISCUSSION

There are two major implications of this study. First, if a learner is processing category exemplars nonanalytically, performance is facilitated when the initial study set is a low-variance sample, with subsequent samples gradually introducing variation. Second, if a learner is processing category exemplars analytically, performance may be facilitated when the initial and subsequent study samples are representative of the category distribution. This interaction is problematic for most schema-abstraction models. Although some models can account for either the first or the second result, there does not seem to be a model that can account for both, that is, for an interaction of information representativeness and processing mode. We will consider these results in the framework of models that can each account for one of the main findings, but not for both.

The advantage of beginning with a low-variance sample and gradually introducing variation can be accounted for by the ACT generalization model (J. R. Anderson, Kline, & Beasley, 1979; Elio & Anderson, 1981). The ACT theory, on which the model is based, proposes a generalization process that is the unconscious, automatic detection of similarities across a set of data. This process creates a representation of those similarities. For example, the Type A items 21113, 21114, 31112, and 41112 would yield generalizations such as 2111-, -1112, -111-, in which a dash indicates a dimension that may take any value. A more detailed application of this model to schema-abstraction tasks can be found in J. R. Anderson et al. (1979) and Elio and Anderson (1981). The important assumption of the model for the present argument is that a feature pattern, either for a specific item or for a generalization, has an associated strength that (1) reflects the number of items it successfully classifies and (2) determines the probability that it will be accessed to classify a given item. When a feature pattern successfully matches an item (i.e., correctly classifies it), the strength of all more general patterns consistent with the specific pattern is also increased. Since this ensures that a generalization will accrue more strength, over time, than any of the specific patterns from which it was formed, eventually these more general feature patterns, rather than patterns corresponding to specific exemplars, will be the basis for item judgments.

A low-variance study sample, in which there is a maximum amount of similarity among items, would be particularly conducive to forming strong category

generalizations. Elio and Anderson (1981) showed that generalizations are more likely to be formed when similar instances occur close together in the study sequence. Although the centered condition's initial sample may have provided a strong set of generalizations about the category's most frequently occurring members, these generalizations would not match members of less frequently occurring member types on subsequent study and test phases. However, contrary to our initial intuitions that centered-condition performance would begin to deteriorate relative to that in the representative condition, it rose on subsequent blocks. It is possible that although the centered condition's second and third samples were less likely to match previously formed generalizations, they were not so disparate as to prevent their refining or strengthening old generalizations. The similarity between adjacent member types made it likely that new items in these samples would both partially fit existing generalizations and give rise to new generalizations. For example, a generalization like 2111- formed from Type A members would not match a Type B member such as 22114. Since the generalization process operates on all feature patterns—both those corresponding to specific exemplars and those corresponding to generalized patterns—encountering an item like 22114 would spawn another pattern such as 2-11- that could then classify a wider range of items. In order for the generalization mechanism to detect such regularities and successfully yield more general feature patterns, there must be a minimal amount of discrepancy between previously formed generalizations and new item patterns.³ The gradual incorporation of variation in the centered condition would be particularly conducive to such processes.

A different kind of model is required to account for the relative performance of rule-condition subjects given representative and centered samples. We assume that subjects are in a more explicit hypothesis-testing mode under these circumstances. It has been shown (Levine, 1966; Trabasso & Bower, 1966) that subjects have rather poor memory for the instances on which their hypotheses are based. As noted earlier, the initial rules formulated from a centered, low-variance sample would be inappropriate for a considerable number of category members. If subjects were not storing complete instance-specific information, they would not have the information necessary to revise their rules when they began to fail. Any revisions they did make may have been governed by their hypotheses about which dimensions were relevant and which were irrelevant. In the centered condition, these hypotheses would be biased toward the Type A members by the initial low-variance sample. Although subjects in the representative condition may not have been able to formulate a set of rules that completely described the category structure after one sample, at least they were not biased to (1) disregard certain features and values as irrelevant or (2) believe certain features or value patterns were perfectly

predictive of one category or another. This may account for their improved performance relative to that of subjects in the centered condition in analytic mode.

Since we have claimed that a generalization mechanism adjusted inadequate generalizations in Experiment 1, but that rule revision in analytic mode would be unsuccessful, it is important to clarify the distinction between a generalization and a rule. Both can be described as feature patterns with various contingencies among dimensions and values. However, we claim the difference is what and how information is processed. The generalization process is posited to be an automatic and unconscious detection of co-occurring feature patterns. A rule is taken to be the product of a conscious, active hypothesis-testing effort. The examples given for Experiment 2's protocols are good illustrations of such rules. The information on which the generalization mechanism works is the encoded feature patterns corresponding to specific exemplars. The more accurately these encoded feature patterns correspond to studied items, the more useful generalizations abstracted from them will be. When the subject is actively formulating rules and hypotheses, we assume that he or she is encoding only a subset of stimulus attributes, presumably those attributes currently considered to be relevant to category membership. This stored exemplar information could be regarded as somewhat impoverished. Thus, we do not propose that a generalization mechanism does not operate when subjects are formulating rules, but that the stored information on which the mechanism works is considerably different.

It is important to consider how instance-based models that use interitem similarity to account for schema-abstraction performance apply to the present data. According to an instance-based model such as Medin and Schaffer's (1978) model, classification of items, both old and new, is based on their similarity to stored exemplars. Instance-based models posit that no higher order information is abstracted about the category during learning. Medin (Note 2) argued that one very attractive feature of such models, relative to a prototype approach, is the ease with which old information can be updated when new information is acquired. It seems this argument would predict that item order should have no ultimate effect on performance, although the present study indicates otherwise. Medin and Schaffer's model does, however, include a selective-attention mechanism that enables the model to predict that initial items could influence the dimensions to which a subject attends and hence how subsequent exemplars are encoded. Medin and Smith (1981) demonstrated how this component of the context model could predict transfer performance under different instructional or strategic sets that presumably bias learners to encode different aspects of the stimuli. With only variations in similarity parameters for each attribute, the context model was able to account for the superior performance of subjects given neutral or "look for rules plus exceptions" instruc-

tions relative to subjects given "form an average [i.e., prototype] member for each category" instructions. Thus, the context model might predict poorer performance given an initial low-variance sample, a result the present study did find, but only when subjects were processing the material analytically.

In summary, by presenting category information bit by bit, we were able to identify some factors affecting the evolution of a schema, rather than simply assessing the end product. The present study underscores the importance of the analytic/nonanalytic learning distinction in this sort of learning. Our view is that the analytic/nonanalytic distinction is best regarded as determining what information gets encoded and the inductive processes that act upon the encoded information. In the analytic mode, only selected aspects of exemplar information are encoded and conscious hypothesis testing takes place. In nonanalytic mode, a more unbiased sample of information is encoded, but only simple, automatic generalization processing occurs. Which combination of information and processes occurs will depend on characteristics of the problem, the to-be-learned material, and the learner. These experiments have shown that one important problem characteristic is the order in which the information is acquired.

REFERENCE NOTES

1. Fried, L. S., & Holyoak, K. J. *Induction of category distributions: A framework for classification learning*. (Cognitive Science Tech. Rep. 38). Ann Arbor: University of Michigan, Department of Psychology, March 1982.
2. Medin, D. L. *Levels of categorization and the integration of new experiences*. Paper presented at the meeting of the American Psychological Association, Montreal, September 1980.

REFERENCES

- ANDERSON, J. R., KLINE, P. J., & BEASLEY, C. M. A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1979.
- ANDERSON, N. R. Application of linear-serial model to a personality impression task using serial presentation. *Journal of Personality and Social Psychology*, 1968, **10**, 354-362.
- ASCH, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, **41**, 258-290.
- BARRESI, J., ROBBINS, D., & SHAIN, K. The role of distinctive features in the abstraction of related concepts. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 360-368.
- BROOKS, L. Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Erlbaum, 1978.
- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1956.
- ELIO, R., & ANDERSON, J. R. The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 397-417.
- FRANKS, J. J., & BRANSFORD, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology*, 1971, **90**, 65-74.
- HAYES-ROTH, B., & HAYES-ROTH, F. Concept learning and the

- recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 321-338.
- HOMA, D. Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 407-416.
- HOMA, D., & VOSBURGH, R. Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 101, 322-330.
- HOVLAND, C. L., & WEISS, W. Transmission of information concerning concepts through positive and negative information. *Journal of Experimental Psychology*, 1953, 45, 175-182.
- LEVINE, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, 71, 331-338.
- MEDIN, D. L., & SCHAFFER, M. M. Context theory of classification learning. *Psychological Review*, 1978, 85, 207-238.
- MEDIN, D. L., & SMITH, E. E. Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, 7, 241-253.
- NEUMANN, P. G. An attribute frequency model for the abstraction of prototypes. *Memory & Cognition*, 1974, 2, 241-248.
- POSNER, M. I., & KEELE, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, 77, 353-363.
- REBER, A. S. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 855-863.
- REBER, A. S. Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 88-94.
- REITMAN, J. S., & BOWER, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 1973, 4, 194-206.
- ROSCH, E., & MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, 7, 573-605.
- RUMELHART, D. E., & NORMAN, D. A. Accretion, tuning, and restructuring: Three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, N.J.: Erlbaum, 1978.
- TRABASSO, T., & BOWER, G. H. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*, 1966, 3, 163-173.

NOTES

1. Our use of the term "transfer test" is nonstandard, since by the last block, all items had been seen in a study phase.
2. All Newman-Keuls tests reported here were significant at the .05 level.
3. A more detailed presentation of the ACT model's mechanisms that correspond to these processes can be found in J. R. Anderson et al. (1979).

(Manuscript received July 29, 1982;
revision accepted for publication August 22, 1983.)