

The effect of category learning on sensitivity to within-category correlations

SETH CHIN-PARKER and BRIAN H. ROSS

University of Illinois at Urbana-Champaign, Urbana, Illinois

A salient property of many categories is that they are not just sets of independent features but consist of clusters of correlated features. Although there is much evidence that people are sensitive to between-categories correlations, the evidence about within-category correlations is mixed. Two experiments tested whether the disparities might be due to different learning and test tasks. Subjects learned about categories either by classifying items or by inferring missing features of items. Their knowledge of the correlations was measured with classification, prediction, typicality, and production tests. The inference learners, but not the classification learners, showed sensitivity to the correlations, although different tests were differentially sensitive. These results reconcile some earlier disparities and provide a more complete understanding of people's sensitivities to within-category correlations.

People experience the world not as a confusing and unpredictable wave of sights and sounds, but as a fairly well organized and familiar environment. This coherence is in part due to the categories that they recognize in the world, categories that allow people to take advantage of the knowledge they have gained from experience. Each time a person wants to classify an instance, communicate an idea, make a prediction about some missing feature, or perform any of many other tasks, knowledge of category membership comes into use (Ross & Spalding, 1994). This knowledge includes knowledge of the features of the category members and of how those features relate.

Categories correspond to the rich correlational structure of the world, and people's knowledge reflects that structure (Malt & Smith, 1984; Rosch, 1975), helping people to distinguish between members of different categories (Billman & Knutson, 1996; Medin, Altom, Edelson, & Freko, 1982). Although categories tend to capture between-categories feature correlations (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), it is not clear how sensitive people are to correlational structure when it occurs within category, instead of occurring between categories.

In the case of a between-categories correlation, the co-occurrence of features is predictive of category membership beyond any predictiveness of the features. That is,

although the features involved in the correlation may provide some information about the category membership, the fact that they occur together in members of a particular category adds to this predictiveness. An extreme example of a between-categories correlation is the exclusive-or category structure, where the features are not predictive of category membership but the co-occurrence of the features is perfectly predictive. In contrast, a *within-category correlation* does not add to the predictiveness of the features. Typically, this type of correlation is found in categories that have multiple features that can instantiate a given attribute. The presence of a within-category correlation does provide information about the category members and how their features relate, but it does not add information for determining category membership.

An example of a within-category correlation can be found in the bicycle category. Bicycles typically have one of two types of tires, knobby tires or slick tires, and one of two types of handlebars, a dropped handlebar or a straight handlebar. Although it may not be immediately apparent, after looking at and riding bicycles for some time, one may notice that there is a strong within-category correlation between the type of tire that a bicycle has and its handlebar. Knobby tires accompany a straight handlebar, and slick tires occur with a dropped handlebar; both combinations can be found on what would be considered good members of the bicycle category. Although the classification of an item as a bicycle may make use of the presence of tires and a handlebar, the within-category correlation involving the tires and handlebar does not provide additional evidence for this classification. This is illustrated by the fact that you can have a perfectly good bicycle with a straight handlebar and slick tires; the combination is simply not usual. If people are able to note this type of co-occurrence, it must be due to some information that has been stored about how the features of the

This work was supported by Grant NSF SBR 97-20304 from the National Science Foundation. We thank John Paul Minda, Gregory Murphy, and Andrea Patalano for their helpful comments on an earlier version of this article. Also, we thank Melissa Paulson for her assistance in the data collection. Correspondence should be addressed to S. Chin-Parker, Beckman Institute, University of Illinois, 405 N. Matthews Ave., Urbana, IL 61801 (e-mail: chinpark@s.psych.uiuc.edu).

— Accepted by previous editorial team

category members are related and not just to the fact that category membership has been correctly identified.

An important point to note is that the distinction between within-category correlations and between-categories correlations depends on what level of categorization is considered. The tire/handlebar correlation would become a between-categories correlation if you were to instead consider the different subcategories of bicycles: The co-occurrence of knobby tires and a straight handlebar distinguishes a mountain bike from other types of bicycles. For the example given, and for the experimental materials, we explicitly provide the level of categorization so that a distinction between the two types of correlations can be maintained.

Access to information about within-category correlations is important in a number of ways. First, any knowledge of how features relate to one another allows for more specific predictions to be made about the features of category members. Second, knowledge of within-category correlations reflects more specifically the relational structure of the world. Just as between-categories correlations often help us to distinguish between categories, the within-category correlations can help to distinguish between different subcategories found within a larger category, allowing for a more accurate organization of knowledge. Third, by indicating the presence of relations between features, information about within-category correlations provides a basis for causal or explanatory grounding. All these considerations make it worthwhile to understand our sensitivity to within-category correlations.

Even though knowledge of within-category correlations can be beneficial, it is unclear from previous research whether people are sensitive to this information. Murphy and Wisniewski (1989; see also Kaplan & Murphy, 1999) found no evidence that subjects explicitly store within-category correlations for use in classification other than for cases in which the correlated features are related by prior knowledge. Ahn (1999) also observed a lack of sensitivity to the presence of correlated features within category members when making classification decisions. However, in a study by J. R. Anderson and Fincham (1996), there was clear evidence that subjects were using information about within-category correlations to make predictions following learning. Similarly, subjects in a study by Thomas (1998; see also Ashby & Gott, 1988, Experiment 3) showed that they had access to correlated feature information found within large, overlapping categories comprising simple visual stimuli.

The research into sensitivity to within-category correlations has produced different results, without a clear picture as to why. There are important differences between the studies mentioned with regard to how the categories were learned and how the correlation information was tested. By studying the effect of these factors, we should be able to better understand people's sensitivity to within-category correlations.

Learning

Recently, Yamauchi and Markman (1998, 2000b; Yamauchi, Love, & Markman, 2002) and A. L. Anderson, Ross, and Chin-Parker (2002) have contrasted classification learning with inference learning. These two learning types are similar in that learning is accomplished by the prediction of some missing category information on the basis of the information present. The main difference is that, during classification learning, the category label is predicted and, during inference learning, a missing feature is predicted. Each learning type has a different focus during learning, a focus that might influence the learning of within-category correlations.

In classification learning, the subject learns to distinguish categories. Upon seeing the features of a given exemplar, a decision is made as to which category label best matches the available information. Classification focuses the learner on the information available that is predictive for distinguishing category membership (Tversky, 1977). The presence of within-category correlations could be overlooked in this learning situation, since they do not aid in predicting category membership.

In inference learning, the subject sees an incomplete exemplar, one that includes the category label but not all of the features. The subject learns about the category by predicting a missing feature on the basis of the category information available. Rather than being concerned with multiple categories, as is the case when distinguishing category membership during classification learning, the learner is focused on only one category during the decision process. In this way, the inference task allows the learner to assess how the features relate to one another, along with the category label. This focuses the learner on the internal category structure, possibly making within-category correlations noticeable.

The distinction between classification learning and inference learning may be an important factor in noticing within-category correlations. The study by Murphy and Wisniewski (1989) used classification learning and found no evidence that the subjects were sensitive to the within-category correlations. The learning task used by J. R. Anderson and Fincham (1996) included an inference-like learning component, recreating the features of an exemplar from a generic exemplar following classification, and they found that subjects were sensitive to within-category correlations. A possible explanation of these disparate findings is that classification does not promote the learning of within-category correlations, whereas inference learning supports it. We included both classification learning and inference learning conditions in these experiments in order to examine this possibility.

Tests

Since information about both the features and their relations is assessed when testing for knowledge of within-category correlations, it is important to consider how a given test addresses both those aspects of category knowl-

edge. We consider four tests: a classification test, a feature prediction test, a typicality-rating task, and a production test.

One way to test category knowledge is to measure how accurately a person can classify instances. However, it is unclear whether this would be a sensitive measure of someone's knowledge of within-category correlations. Although an instance that does not preserve a known correlation may not appear to be as good a category member as an item that maintains the correlation, it seems likely that a person would still believe it is a category member, as opposed to being the member of a contrast category. Since the main objective in a classification test is to determine category membership, this test may not be a sensitive measure of knowledge of within-category correlations.

The studies by J. R. Anderson and Fincham (1996) and Thomas (1998) used a prediction test following learning. In a prediction test, the subject predicts a feature on the basis of information about the other features present in a given category member. Although a prediction test can be thought of as similar to a classification test, in that available information guides the selection of a missing piece of information (the missing information being the category label in the case of a classification test and a feature in the case of the prediction test), the test actually requires the use of very different knowledge (Yamauchi & Markman, 2000a). The prediction test focuses on knowledge about the features within a single category, so it is more likely to be a sensitive measure of knowledge of the relations among those features.

A typicality-rating task provides a measure of how "good" a category member is. The subject is told what category to consider when making the rating, so the focus is on one category during the decision process. Malt and Smith (1984) found that the violation of a known correlation within a given exemplar can reduce a person's typicality judgment, but only if the violation is quite salient. The amount of effect that the correlation information has during the typicality-rating task can give an idea of how salient the within-category correlations are to the subjects.

A production test evaluates a subject's ability to recall specified knowledge. In the case of within-category correlations, subjects are asked to recall features that co-occurred in the category members they had seen before. This is a direct test of whether subjects have explicit knowledge of any within-category correlations following learning.

The Present Experiments

The goal of the present experiments was to examine how learning affects people's sensitivity to within-category correlations, and how different tests measure that knowledge. After considering what each learning task entails, it seems likely that the inference learners will be more sensitive to within-category correlations than will the classification learners, because of the different focuses during the two learning tasks. The focus on within-category information by the inference learners should make the

relational structure of the features more apparent, thus highlighting within-category correlations. The focus of the classification learners should not emphasize the within-category correlations, since they are not predictive of category membership. Our understanding of the tests suggest that they should be differentially sensitive measures of the subjects' knowledge of within-category correlations, the classification test likely being the least sensitive measure.

EXPERIMENT 1

The categories developed for Experiment 1 were described as two work projects at a large company. The category members were designated by "employee files" that listed five characteristics of a given employee, along with the work project that the employee had been assigned to. Within these employee files, two of the characteristics were perfectly correlated in their occurrence, these pairs being the within-category correlations of interest.

It is important to understand the stimulus structure in this experiment. We designed the materials so that knowledge of the within-category correlations was not necessary to be successful in either the classification or the inference learning task. The predictive features were perfectly predictive of the category membership, so the correlation did not add to the predictiveness of an exemplar. This meant that the classification learners did not gain from using information about the within-category correlations during their learning task. Similarly, during the inference learning task, the missing feature could be chosen simply on the basis of the category label present. The subject had a choice of two possible features to replace the missing characteristic in the employee file. One of the choices was consistent with the project label provided, and the other choice was found only in members of the contrast category. As with the classification learners, the inference learners did not have to make use of the correlations in order to be successful during their learning task.

Method

Design. The experiment had two between-subjects factors of interest, learning condition (inference learning vs. classification learning) and test ordering (prediction test first vs. double feature classification test first), and two between-subjects counterbalancing variables, the correlated pairing (Pairing A vs. Pairing B) and correlated pair position (Position 2–3 vs. Position 4–5). This design resulted in 16 learning groups.

Subjects. The subjects were 35 undergraduates from the University of Illinois, who participated for either course credit or pay. Two subjects were randomly assigned to each learning group. Three subjects who did not meet the learning criterion were replaced with individuals from the same population.

Materials. The categories were described as two work projects, labeled *Project D* and *Project K*. The subjects learned about the category members through employee files that, in their complete form, indicated the project the employee was assigned to, and the five characteristics of that employee. The specific features used to instantiate each of the characteristics in Experiment 1 can be found

in the Appendix. The first characteristic listed was nonpredictive and indicated the employee's mail preference. The other four characteristics indicated the employee's work experience, work skill, personality, and education. The features specifying the last four characteristics were perfectly predictive of category membership.

An abstract representation of the category structure can be found in Table 1. For each predictive characteristic found in an employee's file, there are four possible feature instantiations. Two of these features were predictive of Project D, whereas the other two were predictive of Project K. Notice that for the Position 2-3 condition, the specific feature found in the second employee characteristic is perfectly correlated with the feature in the third characteristic. The specific correlated pairings can be found in Table 2. The features contained within the fourth and fifth characteristics were not correlated with any other features within the category. The design of the Position 4-5 condition was the same, except that the fourth and fifth characteristics always contained the correlated pairing and the features within the second and third characteristics were not correlated. In each experimental condition, there were 16 employee files total, 8 in each project. All employee files were seen during learning.

The test items in the double feature classification test consisted of pairs of features that came from the same category. Four of the test items contained an *intact correlation*, four test items contained a *broken correlation*, and eight filler test items contained pairs not correlated during study. The intact correlations were the four correlated pairs of features seen during learning (see Table 2 for specific pairings). The broken correlations were never seen during learning; they were made by scrambling the correlated pairs from within the same project. The important thing to remember is that all of the features in these test items were perfectly predictive of category membership. The only way in which the broken correlation test items differed from the intact correlation test items was whether the relational information from the learning items was preserved.

Each prediction test item consisted of one presented feature and two choices: Both of the choices were features from the same category and were possible features for the same characteristic. Since both of the choices occurred equally often as characteristics for the employees of a given project, the only way that one of the choices could be considered better than the other choice was if a relation between that choice and the presented feature was known. For example, a subject that had learned Pairing A for Project K (see Table 2)

would see "engineering degree" as a presented feature. The two choices would be "hard-working" and "organized." Both of these choices were presented equally often in Project K; the only difference was that, in this case, "hard-working" always co-occurred with "engineering degree." It is important to note that although this may seem similar to the inference learning task, the two tasks were qualitatively different. During inference learning, the task was to identify which of two features was consistent with the project assignment of the employee. During the prediction test, the task was to identify which of two features maintained a within-category correlation. There were 16 prediction test items, 8 of which dealt with the correlated features and 8 of which were filler.

The 16 test items in the typicality-rating task were complete employee files. Eight test items had an intact correlation present in the exemplar; half of these test items had features that were all predictive of the category label, and half included one feature (not involved in a correlated pairing) that was predictive of the contrast category. The other 8 test items contained a broken correlation; half had features that were all predictive of the category label, and half had one feature that was predictive of the contrast category.

Procedure. After being randomly assigned to a learning group, the subjects were given verbal instructions; all subsequent instructions and reminders appeared on the computer screen. All learning and testing was done on a Macintosh computer using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993), except for the final questions about the categories and learning, including the paired-features production task, which each subject answered in writing. All the trials within learning blocks and tests were randomly presented.

In the classification learning condition, the subjects saw complete employee files but no project label. The characteristics were listed in the same order in the center of the screen for every subject. The subjects indicated which project they thought the employee was assigned to by pressing the "D" or the "K" key on the keyboard. If the choice was correct, a large, green CORRECT appeared for 2 sec. If the choice was incorrect, a large, red INCORRECT appeared for 2 sec. Following feedback, the subject was shown the complete employee file, including the project label, to study. Study time was self-paced. The learning phase continued for a minimum of three blocks, 16 exemplars per block, until the subject was able to correctly identify 14 of the 16 employee assignments within a block.

In the inference learning condition, the subjects were presented with an incomplete employee file that gave the project label, the nonpredictive characteristic, and three of the four predictive characteristics. The missing characteristic was replaced by a series of question marks. The project label appeared first on the screen for 1 sec, and then the characteristics were listed below it. The files for employees from Project D were centered on the right half of the screen, and the files for employees from Project K were centered on the left half of the screen.¹ Two possible features for the missing characteristic, a choice *a* and a choice *b*, were presented one above the other in the lower center part of the screen. The subjects indicated their choices by using the "A" key or the "B" key on the keyboard. The feedback, self-paced study time, and learning criterion were the same as those in the classification condition.

Following learning, the subjects were tested without feedback. The testing for each subject was the same, except for a balancing of the order of the double feature classification test and the prediction test between subjects. These two tests were always presented first and second in the test order, followed by the typicality-rating task and the paired-features production test. During the classification and prediction tests, following each test item, the subject was asked to give a confidence rating of his or her decision, using a scale of 1 (*guessing*) to 7 (*very confident*).

In the double feature classification test, a pair of features from the same category was presented in the center of the screen. The subject indicated which project an employee with those features belonged to by pressing either the "D" or the "K" key on the keyboard.

Table 1
Design of Categories in Experiment 1

Project K					Project D				
C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
Position 2-3									
np	1	1	1	1	np	3	3	3	3
np	1	1	1	2	np	3	3	3	4
np	1	1	2	1	np	3	3	4	3
np	1	1	2	2	np	3	3	4	4
np	2	2	1	1	np	4	4	3	3
np	2	2	1	2	np	4	4	3	4
np	2	2	2	1	np	4	4	4	3
np	2	2	2	2	np	4	4	4	4
Position 4-5									
np	1	1	1	1	np	3	3	3	3
np	1	2	1	1	np	3	4	3	3
np	2	1	1	1	np	4	3	3	3
np	2	2	1	1	np	4	4	3	3
np	1	1	2	2	np	3	3	4	4
np	1	2	2	2	np	3	4	4	4
np	2	1	2	2	np	4	3	4	4
np	2	2	2	2	np	4	4	4	4

Note—C, employee characteristic; np, nonpredictive characteristic (two possible features which occur equally often in each category).

Table 2
Correlated Pairs for the Experimental Materials

Pairing	Project K	Project D
Correlated Pairs for Experiment 1		
A	engineering degree/hard-working math degree/organized	physics degree/sociable computer degree/analytic
B	marketing experience/speaking skills sales experience/artistic skills	advertising experience/research skills public relations experience/writing skills
Correlated Pairs for Experiment 2		
A	psychology degree/racquetball club geology degree/volleyball club	sociology degree/bowling club anthropology degree/running club
B	available mornings/resourceful available afternoons/hard-working	available evenings/punctual available weekends/organized

During the prediction test, one feature was presented in the center of the screen and two choices, labeled *a* and *b*, were presented below and to each side of that feature. The subject pressed either the “A” or the “B” key on the keyboard to indicate which choice feature they thought belonged with the presented feature.

Each item in the typicality test was a complete employee file presented in the center of the screen, and the subjects were asked to rate how typical that employee was of the project indicated. The items were rated on a scale from 1 (*not typical*) to 7 (*very typical*).

Following the test phase on the computer, the subjects were asked how they approached their learning task and about the projects they had learned. The questions of particular interest concerned the paired-features production test. The subjects were asked, “Did you notice any characteristics in Project D employees that occurred together?” and the same for Project K.

Results and Discussion

All the results for Experiment 1 are collapsed across the test ordering factor and the two counterbalancing variables so that the main effects of learning type and test type can be clearly assessed. For all test data collected, the accuracy measurements are reported as the group mean proportion correct. The confidence ratings for test items incorrectly answered were multiplied by -1 . This set the confidence rating for someone responding at chance at 0. The means for each of the test results for Experiment 1 can be found in Table 3. The data for each test were analyzed with *t* tests focused on the three questions of interest. First, did the classification learners show sensitivity to the within-category correlation? Second, did the inference learners show sensitivity? Third, did the inference learners show greater sensitivity than the classification learners?

Study. There was a difference between the inference learners and classification learners in the number of blocks

needed to reach the learning criterion. The classification learners reached the learning criterion after an average of 3.63 blocks ($SD = 0.96$), whereas the inference learners reached that point after 5.06 blocks [$SD = 2.35$; $t(30) = 2.27, p < .05$]. All 3 subjects who did not meet the learning criterion were from the inference condition.

Prediction test. The prediction test items of interest had a presented feature that had been part of a correlated pair during study. Without information about the correlations, the subjects should choose the correlated feature on .5 of these items, and their confidence ratings should have a mean of 0. The results of the prediction test showed a significant difference between the learning groups. The classification learners were not different from chance in their accuracy ($M = .49, SD = .22$) and their confidence [$M = -0.20, SD = 2.16$; both $t_s(15) < 1$]. The inference learners showed that they knew the correlations well both in their accuracy measurements [$M = .71, SD = .27$; $t(15) = 3.18, p < .01$], and in their confidence ratings [$M = 2.59, SD = 3.35$; $t(15) = 3.10, p < .01$]. A comparison of the means for each group reveals a significant difference between the classification and inference learners, both for accuracy [$t(30) = 2.55, p < .05$] and for confidence [$t(30) = 2.80, p < .01$].

Double feature classification test. The double feature classification test items of interest contained either an intact correlation or a broken correlation. If knowledge of within-category correlations plays a role in classification, one would expect the intact pairs to be more accurately classified, and with more confidence, than the broken correlations. The accuracy difference score for each subject was calculated by subtracting the proportion correct with the broken correlations from the pro-

Table 3
Summary of Results From Experiment 1

Learning Task	Prediction Test		Double Feature Classification Test			Typicality-Rating Task (Typicality Rating)			Paired-Features Production Test (Number of Correct Correlations Generated)			
	Proportion Correct	Confidence Rating	Intact	Broken	Diff	Intact	Broken	Diff	Intact	Broken	Diff	
Classification	.49	-0.20	.70	.67	.03	2.50	2.20	0.30	4.95	4.82	0.13	-0.44
Inference	.71	2.59	.77	.67	.09	3.06	2.31	0.75	5.03	4.20	0.83	1.38

Note—All results reported are group means. Intact, intact correlation; broken, broken correlation; diff, difference between intact and broken. Confidence rating means have a range of -7 to 7 . Typicality rating means have a range of 1 to 7 .

portion correct with the intact correlations. The confidence difference scores were calculated analogously. The classification learners' accuracy showed a non-significant difference between the intact and the broken test items [$M = .03$, $SD = 0.22$; $t(15) < 1$]. The inference learners' accuracy also showed a nonsignificant difference, although the effect was marginal [$M = .09$, $SD = .20$; $t(15) = 1.86$, $p < .10$]. The difference in confidence was nonsignificant for both the classification learners [$M = 0.30$, $SD = 1.69$; $t(15) < 1$] and the inference learners [$M = 0.75$, $SD = 1.89$; $t(15) = 1.56$, $p < .15$]. A comparison of the two learning groups shows no difference between their intact/broken differences in either accuracy or confidence [both $t(30) < 1$]. None of the comparisons of interest showed an effect of the within-category correlations on classification performance, although the inference learners did show a marginal effect for accuracy.

Typicality-rating task. To assess the effect of the correlation information, the average rating for test items containing a broken correlation was subtracted from the average rating for the test items containing an intact correlation for each individual. The inference learners were sensitive to the correlations, significantly more so than the classification learners. The average typicality difference for the classification learners ($M = 0.13$, $SD = 0.65$) was not significantly different from zero [$t(15) < 1$], whereas the difference for the inference learners ($M = 0.83$, $SD = 0.90$) was different from zero [$t(30) = 3.68$, $p < .01$]. The inference learners showed a significantly greater difference than did the classification learners [$t(30) = 2.54$, $p < .05$].

Paired-feature production test. The difference between the number of pairs written containing an intact correlation and those containing a broken correlation was determined for each subject. The classification learners were unable to generate any of the correlated pairs seen during study, but some subjects did generate some pairs that contained a broken correlation ($M = -0.44$, $SD = 0.63$). The inference subjects generated more intact correlations than broken correlations ($M = 1.38$, $SD = 1.67$). A comparison of the group means shows a significant difference [$t(30) = 4.07$, $p < .001$] between the two learning types. Once again, it is evident that the inference learners were more able to draw upon information about the within-category correlations.

Summary. A clear difference between the learning types was seen in this experiment. In three of the four tests, the double feature classification test being the one exception, the inference learners showed that they were able to access information about the within-category correlations more reliably than the classification learners.

EXPERIMENT 2

Experiment 2 was designed to replicate and extend the major findings of Experiment 1. The primary extension of Experiment 2 was to address whether the inference learners' sensitivity to the within-category correlations required the features in the correlated pairs to be adjacent,

as in the prior experiment. It was possible that the inference learners depended on the spatial contiguity of the correlated features in the display. A new inference condition was included in this experiment to examine that possibility.

There were three learning conditions, one classification condition and two inference conditions. The design of the classification condition was identical to that in Experiment 1. The *inference-near* condition was identical to the inference condition in Experiment 1; the correlated pairs occurred as adjacent characteristics. The *inference-split* condition used the same correlated pairings, but they appeared as characteristics in the employee files that were separated by another characteristic. The purpose of the inference-split condition was to determine whether people would be sensitive to the correlations if the features were not in adjacent positions in the list of employee characteristics. We did not include a classification condition in which the correlated pairing was split, because the classification learners did not show sensitivity to the within-category correlations even when they were adjacent.

Method

Design. There were three learning conditions in this experiment. The classification condition was identical to the design in Experiment 1. The *inference-near* condition was identical to the inference condition run in Experiment 1. The *inference-split* condition used the same correlated pairings as the classification and *inference-near* conditions, but the two correlated features were separated by another characteristic in the employee file. The position of the pairing in the *inference-split* condition was balanced across subjects, appearing in the second and fourth positions or the third and fifth positions.

In addition to the three learning conditions, there were a number of counterbalancing variables across subjects: two possible pairings that the subject could be exposed to during learning, two possible positions within the employee file where the pairing could occur, and two possible test orderings (prediction test first vs. double feature classification test first). These manipulations resulted in 24 learning groups overall.

Subjects. The subjects were 56 undergraduates from the University of Illinois, who participated for either course credit or pay and were randomly assigned to a learning group. The data for 1 subject were lost owing to a computer malfunction. Seven subjects did not meet the learning criterion and were replaced.

Materials. All the materials for the learning and testing during this experiment were presented as in Experiment 1. New features within the employee files were chosen to better differentiate the features and lessen the possibility that the subjects could construct ad hoc relations between features during learning (see the Appendix). The correlated pairings used in Experiment 2 are presented in Table 2.

Procedure. The procedures followed for this experiment were identical to those in Experiment 1.

Results and Discussion

The experiment had two comparisons of interest. The first was a replication of Experiment 1, contrasting the classification and the *inference-near* conditions. The second comparison was between the *inference-near* and the *inference-split* conditions, to determine whether the correlated features had to be adjacent in order for an inference learner to show sensitivity to the correlations.

Table 4
Summary of Results From Experiment 2

Learning Task	Prediction Test		Double Feature Classification Test						Typicality-Rating Task (Typicality Rating)			Paired-Features Production Test (Number of Correct Correlations Generated)
	Proportion Correct	Confidence Rating	Proportion Correct			Confidence Rating			Intact	Broken	Diff	
			Intact	Broken	Diff	Intact	Broken	Diff				
Classification	.47	−0.25	.83	.86	−.03	4.00	4.31	−0.31	5.07	4.91	.16	−0.19
Inference-near	.81	3.82	.91	.83	.08	4.97	3.61	1.36	5.12	4.57	.55	1.31
Inference-split	.70	2.45	.91	.80	.11	4.89	3.52	1.38	5.23	4.29	.94	0.44

Note—All results reported are group means. Intact, intact correlation; broken, broken correlation; diff, difference between intact and broken. Confidence rating means have a range of −7 to 7. Typicality rating means have a range of 1 to 7.

The results for the tests in Experiment 2 can be found in Table 4.

Study. There was a marginal difference between classification learners ($M = 3.06$, $SD = 0.25$) and inference-near learners ($M = 3.88$, $SD = 1.75$) in the number of study blocks needed to reach the learning criteria [$t(30) = 1.84$, $p < .10$]. A comparison of the two inference conditions shows that the subjects in the inference-near condition did not differ overall from the subjects in the inference-split condition [$M = 4.88$, $SD = 2.39$; $t(30) = 1.35$, $p < .20$]. All of the 7 subjects who did not reach the learning criterion were in the inference conditions: 3 of these subjects were in the inference-near condition, and 4 were in the inference-split condition.

Prediction test. In the prediction test, both of the inference groups showed sensitivity to the within-category correlations, whereas once again the classification group did not.² The performance of the classification subjects on the prediction test was at chance, with regard to both their accuracy ($M = .47$, $SD = .28$) and their confidence scores [$M = -0.25$, $SD = 2.40$; both $t_s(15) < 1$]. The inference-near subjects performed reliably above chance in their accuracy [$M = .81$, $SD = .26$; $t(15) = 4.84$, $p < .001$] and their confidence score [$M = 3.82$, $SD = 3.16$; $t(15) = 4.83$, $p < .001$]. The inference-split subjects showed a similar effect for both their accuracy [$M = .70$, $SD = .30$; $t(15) = 2.58$, $p < .05$] and their confidence [$M = 2.45$, $SD = 3.31$; $t(15) = 2.96$, $p < .01$].

A comparison of the classification learners and the inference-near learners revealed a significant difference in their accuracy [$t(30) = 3.59$, $p < .01$] and confidence ratings [$t(30) = 4.10$, $p < .001$]. There was no significant difference between the inference-near and the inference-split conditions in their accuracy [$t(30) = 1.18$, $p < .25$] or confidence [$t(30) = 1.20$, $p < .25$]. The first comparison again supports the idea that learning has an effect on the subject's ability to later use information about within-category correlations seen during study. The second comparison undermines the contention that the spatial contiguity of the correlated features would affect the inference learners' sensitivity to within-category correlations.

Double feature classification test. As in Experiment 1, the classification learners showed no effect of the correlations in either accuracy ($M = -.03$, $SD = .18$) or confidence [$M = -0.31$, $SD = 1.28$; both $t_s(15) < 1$]. The inference-near learners did not show an effect of the correlation information on their classification accuracy

[$M = .08$, $SD = .25$; $t(15) = 1.23$, $p < .25$] but did show a difference on their confidence [$M = 1.36$, $SD = 2.39$; $t(15) = 2.28$, $p < .05$]. The inference-split subjects showed significant effects both in accuracy [$M = .11$, $SD = .16$; $t(15) = 2.78$, $p < .05$] and in confidence [$M = 1.38$, $SD = 1.72$; $t(15) = 3.20$, $p < .01$]. A comparison of the classification group and the inference-near group showed a nonsignificant difference in accuracy [$t(30) = 1.41$, $p < .20$] but a significant difference in confidence [$t(30) = 2.47$, $p < .05$]. The same comparisons of the two inference groups found no differences in either accuracy or confidence [both $t_s(30) < 1$].

Typicality-rating task. The typicality-rating task proved to be a less sensitive measure than in Experiment 1, although the general trends found in the results were similar. For classification learners, the difference between the mean typicality ratings of exemplars containing an intact correlation and exemplars containing a broken correlation ($M = 0.16$, $SD = 0.48$) was not different from zero [$t(15) = 1.37$, $p < .20$], whereas both of the inference learning groups had differences that were marginally different from zero [inference-near, $M = 0.55$, $SD = 1.17$, $t(15) = 1.87$, $p < .10$; inference-split, $M = 0.94$, $SD = 1.88$, $t(15) = 2.00$, $p < .10$]. There was no significant difference between the means of the classification learners and the inference-near learners [$t(30) = 1.21$, $p < .25$] or the means of the inference-near learners and the inference-split learners [$t(30) < 1$].

Paired-features production test. The subjects' responses to the final questions were analyzed in the same fashion as in Experiment 1. The difference between the numbers of intact and broken correlations generated by the classification learners ($M = -0.19$, $SD = 0.83$) was less than that for the inference-near learners [$M = 1.31$, $SD = 2.77$; $t(30) = 2.07$, $p < .05$], replicating the result from Experiment 1. The inference-split learners ($M = 0.44$, $SD = 2.28$) were not significantly different, in their ability to recall correlated pairs, from the inference-near learners [$t(30) < 1$].

Summary. The results of the classification learners and the inference-near learners were very similar to those found in Experiment 1. The classification learners showed no sensitivity to the correlation information in any of the tests. The inference learners showed reliable sensitivity to correlation information in the prediction test and the paired-features production test, as in Experiment 1. The inference-near learners did not show any ef-

fect for their accuracy in the double feature classification test, but they did show an effect for their confidence during the test. These results suggest that the information about the within-category correlations is salient enough for the inference learners that it can have a small effect on classification performance. For the inference-near learners, the typicality-rating task proved to be less sensitive in this experiment, indicating that it may not be especially reliable in measuring knowledge of within-category correlations, a point made by Malt and Smith (1984).

The comparison of the inference-near and the inference-split conditions revealed that the positioning of the correlated pair, whether it occurred with adjacent features or not, had no significant effect. However, the inference-near learners did show some advantage, although not statistically reliable, on the most sensitive measures, the prediction test and the paired-features production test.

GENERAL DISCUSSION

These experiments were designed to assess whether subjects are sensitive to within-category correlations and how that sensitivity is affected by the learning task. A secondary consideration was to evaluate several tests of category knowledge with regard to how sensitive they are to knowledge of the correlated features. Across the experiments, the category learning task was critical; subjects learning through the inference of missing features reliably picked up on the within-category correlations, whereas subjects learning through the classification of exemplars did not. The tests of category knowledge were differentially sensitive to knowledge of these correlations. The classification test and the typicality-rating task were not consistently sensitive measures of the subjects' knowledge of the within-category correlations, whereas the prediction test and the production test consistently were.

Relation to Previous Research

The results of these two experiments help to explain why there are some incompatible findings in the literature on people's sensitivity to within-category correlations. Previous research had used different combinations of learning and tests, which here were shown to affect the learner's ability to capture and make use of information about within-category correlations.

Murphy and Wisniewski (1989) found no evidence that the subjects were using information about within-category correlations. Since the subjects in their experiments learned by classification and then were tested with classification tests, their results correspond to what we found in our experiments. Other studies that found a lack of sensitivity to within-category correlations (Ahn, 1999; Kaplan & Murphy, 1999) also used classification measures to assess subjects' knowledge.

In a study by J. R. Anderson and Fincham (1996), subjects categorized examples of irises. The learning task in the second and third experiments was interesting in that, following a classification judgment, the subjects had to

recreate the iris example they had just seen, essentially inferring the features of a specific flower from a generic example. This learning was followed by a prediction test. A focus on the feature relations during learning, coupled with the prediction test following learning, explains why these experiments were successful in illustrating the subjects' sensitivity to within-category correlations.

The differences between learning and test are not sufficient to explain all of the earlier findings: Thomas (1998) found that classification learning led to sensitivity to within-category correlations on a prediction test, a result that conflicts with both of the experiments presented here. The subjects in the Thomas study learned to classify visual stimuli composed of only two continuous dimensions, circle size and line angle. After receiving feedback while classifying over 2,000 exemplars, some of the subjects showed sensitivity to the within-category correlation information on a prediction test. Although this finding suggests that some caution is in order for our interpretation, the simplicity of the materials and the massive amount of training also suggests that there should be some caution in claiming that people are sensitive to within-category correlations following most classification learning.

One aspect of the previous research that has not been addressed is the effect of the type of stimuli. The experiments that found sensitivity to within-category correlations used visual stimuli: pictures of irises (J. R. Anderson & Fincham, 1996) and circle and line combinations (Thomas, 1998). The studies that did not find this sensitivity used stimuli that were feature lists (Murphy & Wisniewski, 1989). Rather than presenting a complete image of a category member, certain features were listed for each category member. There are important differences in how subjects process these different types of stimuli. However, in the present experiments, the inference learners showed sensitivity to the within-category correlations even though the stimuli were feature lists, so this factor is unlikely to be a crucial one.

An Effect of Learning

Recently, there has been an attempt to evaluate the effects of classification learning and inference learning (A. L. Anderson et al., 2002; Yamauchi et al., 2002; Yamauchi & Markman, 1998, 2000a, 2000b). The results of these experiments add to that body of knowledge and support the general view that, during classification learning, the focus is on what differentiates the categories and, during inference learning, the focus is on what binds a category together.

The main purpose of classification is to distinguish between categories, so classification learning leads to an emphasis on information that helps in distinguishing category membership (Yamauchi & Markman, 1998). This focus can result in missing other information, such as information about within-category correlations, that does not aid in this task. It is important to remember that classification learning alone is not the usual way in which

people learn new categories. Classification provides a means for sorting and accessing knowledge about a category for a variety of tasks, such as prediction, explanation, and problem solving. Although people do show some sensitivity to within-category correlations found in real-world categories (Malt & Smith, 1984), the sensitivity may not have arisen from the classification experience. For instance, Murphy and Wisniewski (1989; also Murphy & Medin, 1985) have pointed out that a theory can guide recognition of correlated features.

The main purpose of inference learning is not to distinguish category membership, but to be able to infer a missing feature for a classified item. Thus, the focus of the inference learners is on the within-category structure, how the category label and features relate to one another. The fact that inference learners were sensitive to the within-category correlations indicates that there is something about the process involved in inference learning that promotes the recognition of this type of relational information. Yamauchi and Markman (2000b) proposed that the inference learning process depends on the comparison of members within a category, and this continual comparison can highlight a commonality found across the category members. In this situation, the correlations would become apparent as the comparison highlighted the fact that certain features occurred together. An extension of this idea provides an account of how subcategories can be developed. As a relation between correlated features becomes noticed, there would be some restructuring of the category knowledge so that this within-category correlation can be reflected. Within a given category, the knowledge would be organized to reflect the fact that certain features co-occur, and this organization would necessarily differentiate between instances that had different correlated features. In this manner, the noticing of within-category correlations by the inference learners can lead to the formation of subcategories.

The results of the inference learners are similar to those found in studies with very different learning tasks but a similar focus during the learning. In several unsupervised learning experiments (Billman & Knutson, 1996; Kersten & Billman, 1997), subjects were found to be sensitive to the correlational structure within a group of instances. By taking the focus away from category membership, as was done in the inference learning condition in these experiments, the subjects looked for information that would allow them to develop some coherent knowledge of the category.

Even though there was a clear effect of learning shown in these experiments, some caution is necessary. These results do not imply that classification learning never leads to sensitivity to within-category correlations. Simple stimuli (as in Thomas, 1998), obvious causal or theoretical relations between certain features, or other factors may make it possible for someone classifying to learn this type of within-category relation. Likewise, just because someone learns about a category through inference does not mean that he or she will necessarily be

sensitive to within-category correlations. Complex stimuli or some methods of presentation may counter the effect of inference learning. These experiments were not designed or intended to establish the boundaries of when and how people are sensitive to within-category correlations. Rather, these experiments were intended to clarify disparate results found in previous studies dealing with within-category correlations.

Closing Thoughts

The driving force behind these experiments was to evaluate the sensitivity of people to within-category correlations during category learning. Murphy and Wisniewski (1989) had proposed that prior knowledge constrained the ability to notice and learn within-category correlations. Our results suggest that the learning task, in addition to a knowledge constraint, influences whether feature co-occurrence can be recognized and utilized in establishing the internal structure of a category.

This issue is important because of what these feature co-occurrences provide. As has been noted by Malt and Smith (1984), the natural world is a source of rich and systematic feature correlations. Recognizing these feature relations is the first step to understanding the reasons they exist. Medin (1989) discusses the importance of recognizing that superficial feature structure (i.e., similarity) can be an invaluable tool for pointing toward the deeper relational structures that are essential for a knowledge-based system of categorization. Once we have recognized the fact that features are correlated, we can draw upon these correlations to guide our development of causal theories, allowing for an even richer and more complete representation of the category.

REFERENCES

- AHN, W.-K. (1999). Effect of causal structure on category construction. *Memory & Cognition*, *27*, 1008-1023.
- ANDERSON, A. L., ROSS, B. H., & CHIN-PARKER, S. (2002). A further investigation of category learning by inference. *Memory & Cognition*, *30*, 119-128.
- ANDERSON, J. R., & FINCHAM, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 259-277.
- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 33-53.
- BILLMAN, D., & KNUTSON, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 458-475.
- COHEN, J. D., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257-271.
- KAPLAN, A. S., & MURPHY, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition*, *27*, 699-712.
- KERSTEN, A. W., & BILLMAN, D. (1997). Event category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 638-658.
- MALT, B. C., & SMITH, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning & Verbal Behavior*, *23*, 250-269.

- MEDIN, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, **44**, 1469-1481.
- MEDIN, D. L., ALTOM, M. W., EDELSON, S. M., & FREKO, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 37-50.
- MURPHY, G. L., & MEDIN, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.
- MURPHY, G. L., & WISNIEWSKI, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and application* (pp. 23-45). Englewood Cliffs, NJ: Prentice-Hall.
- ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.
- ROSCH, E., MERVIS, C. B., GRAY, W. D., JOHNSON, D. M., & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.
- ROSS, B. H., & SPALDING, T. L. (1994). Concepts and categories. In R. J. Sternberg (Ed.), *Handbook of perception and cognition: Vol. 12. Thinking and problem solving* (pp. 119-148). San Diego: Academic Press.
- THOMAS, R. D. (1998). Learning correlations in categorization tasks using large, ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 119-143.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.
- YAMAUCHI, T., LOVE, B. C., & MARKMAN, A. B. (2002). Learning non-linearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 585-593.
- YAMAUCHI, T., & MARKMAN, A. B. (1998). Category learning by inference and classification. *Journal of Memory & Language*, **39**, 124-148.
- YAMAUCHI, T., & MARKMAN, A. B. (2000a). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 776-795.
- YAMAUCHI, T., & MARKMAN, A. B. (2000b). Learning categories composed of varying instances: The effect of classification, inference, and structural alignment. *Memory & Cognition*, **28**, 64-78.

NOTES

1. During inference learning, it was necessary to emphasize the category membership of each exemplar (by presenting the category label first and having the employee files off-center) so that, following learning, the inference, subjects would be able to classify at an acceptable level. When piloting designs, we found that when the learning exemplars were presented without the salient category differences, the inference learners were able to predict missing features but were classifying features at chance. It seemed that they learned the within-category structure but had not learned the connection between the feature information and the category label. Since we were interested in the learning effect on classification performance, we made the category membership more salient during learning.

2. Although the classification learners in these two experiments show no evidence of any effect on the prediction test, one pilot experiment did show a significant effect of .60. This experiment is not included here because several experimental variables were not balanced across subjects and there was concern that the correlated pairings were poorly designed. However, even though the effect was significant, half of the difference from chance was due to only 2 of the 16 subjects. Without those 2 subjects, the effect was only .55 and not significant [$t(13) = 1.88, p < .10$]. On the basis of these issues, we do believe the evidence points to little or no sensitivity to within-category correlations for classification learners, but we address the issue further in the General Discussion section, when examining the results of Thomas (1998).

APPENDIX

Features in Experiment 1

Mail Preference (nonpredictive):

- Prefers e-mail
- Prefers postal mail

Education:

- Computer degree
- Engineering degree
- Math degree
- Physics degree

Experience:

- Advertising experience
- Marketing experience
- Public relations experience
- Sales experience

Work Skill:

- Artistic skill
- Research skill
- Speaking skill
- Writing skill

Personality:

- Analytic
- Hard-working
- Organized
- Sociable

Features in Experiment 2

Communication preference (nonpredictive):

- Prefers telephone
- Prefers meetings

Company club:

- Bowling club
- Racquetball club
- Running club
- Volleyball club

Personality:

- Hard-working
- Organized
- Punctual
- Resourceful

Education:

- Anthropology degree
- Geology degree
- Psychology degree
- Sociology degree

Availability:

- Available mornings
- Available afternoons
- Available evenings
- Available weekends