Category inference as a function of correlational structure, category discriminability, and number of available cues

Matthew E. Lancaster · Ryan Shelhamer · Donald Homa

Abstract Two experiments investigated category inference when categories were composed of correlated or uncorrelated dimensions and the categories overlapped minimally or moderately. When the categories minimally overlapped, the dimensions were strongly correlated with the category label. Following a classification learning phase, subsequent transfer required the selection of either a category label or a feature when one, two, or three features were missing. Experiments 1 and 2 differed primarily in the number of learning blocks prior to transfer. In each experiment, the inference of the category label or category feature was influenced by both dimensional and category correlations, as well as their interaction. The number of cues available at test impacted performance more when the dimensional correlations were zero and category overlap was high. However, a minimal number of cues were sufficient to produce high levels of inference when the dimensions were highly correlated; additional cues had a positive but reduced impact, even when overlap was high. Subjects were generally more accurate in inferring the category label than a category feature regardless of dimensional correlation, category overlap, or number of cues available at test. Whether the category label functioned as a special feature or not was critically dependent upon these embedded correlations, with feature inference driven more strongly by dimensional correlations.

Keywords Categorization · Concepts · Inductive reasoning · Expertise · Memory

M. E. Lancaster · R. Shelhamer · D. Homa (✉)
Department of Psychology, Arizona State University,
Tempe, AZ 85287, USA
e-mail: donhoma@asu.edu

The experimental study of human categories has an extensive history, beginning with the seminal study of Hull (1920) and continuing today into the identification of variables critical to the shaping of concepts and the development of formal, quantitative models of classification.

Hull introduced the classification paradigm that dominates most current research today. In this paradigm, the subject initially assigns a number of patterns into designated categories, followed by a transfer test containing old and new instances. By manipulation of variables in the learning phase and then evaluating transfer performance, Hull was able to draw a number of conclusions about the learning and representation of concepts; for example, concepts were learned more rapidly in the order from simple to complex rather than the reverse, transfer was better following learning of many patterns shown infrequently rather than a few patterns presented numerous times, and so forth.

However, categories provide functions above and beyond classification. Bruner, Goodnow, and Austin (1966) summarized a number of additional utilities of categories: Once learned, they permit generalization to novel instances, thereby reducing the need for new learning; they simplify the incredible complexity of the environment into a manageable set of units, thereby facilitating a host of cognitive functions, including logical reasoning and communication; they are adaptive so that harmful or threatening stimuli can be responded to appropriately; and they permit inferences of hidden or obscure attributes when full stimulus information is lacking.

The last property—the inference of missing or unavailable information—is the focus of the present study. Inference is likely to arise whenever less than complete information is available. For example, a physician might suspect a disease on the basis of the presentation of initial symptoms. Confirmation or increased confidence in diagnosis arises when identification of other characteristics likely

associated with that disease is found. Swets, Dawes, and Monahan (2000) provided numerous examples of precisely this logic, in which accuracy of a diagnosis, such as evidence of prostate cancer, is increasingly improved when additional tests consistent with that disease are obtained.

The investigation of attribute inference has received considerably less attention than has classification, with research primarily focused on whether subjects are sensitive to the internal correlational structure of the categories following learning and whether the category label functions as a special feature. Medin, Altom, Edelson, and Freko (1982) had subjects initially study cases of a fictitious disease defined by multiple dimensions, two of which were perfectly correlated or uncorrelated. On the subsequent transfer test, subjects were provided with two test pairs, one of which preserved the correlation. In general, subjects selected, as a member of the disease, the stimulus that preserved the correlation, demonstrating that feature combinations were either stored or computed at the time of decision. Lassaline and Murphy (1996) initially presented stimuli for inspection, followed by questions about feature inference or frequency; both groups then were allowed to sort the original set into two categories that seemed most natural. Interestingly, the prior inference task was more likely to produce sorting that mirrored a family resemblance structure; the prior frequency task generally produced categories in which a single dimension was used to discriminate between the two categories.

Yamauchi and Markman (1998) addressed inference in a novel paradigm that required the subject to identify a feature value or the category label. They noted that the two tasks were formally identical if the category label functioned as simply another feature. For example, stimulus i might be represented schematically as $\{f_1, f_2, f_3, \ldots f_n\}$, where the initial feature (f_1) is the category label and features f_2–f_n are the presented attributes. When classification is studied, the subject is typically given a set such as $\{?\ f_2, f_3, \ldots f_n\}$ and is asked to identify f_1, the category label; when given the set $\{f_1, f_2, ?, \ldots f_n\}$, the label and all features but one are provided, and the subject is asked to infer the omitted feature.[1] In their task, the subject learned the categories either by inference training or by the traditional classification method. That is, when the subject was provided the full feature set and asked to classify the stimulus, the task was a standard classification task; that is, the category label had to be inferred. In contrast, when the subject was asked to

identify a missing feature, given the category label and the remaining features but one, the task became a feature inference task.

Inference training led to better performance on inference transfer, and classification training produced better performance on classification transfer. In addition, inference training led to a higher probability of inferring a prototypical value for a missing feature than did classification training. As a result, Yamauchi and Markman (1998) asserted that inference training and classification training generated different categorical representations.

Studies following Yamauchi and Markman (1998) have generally found that inference training generated knowledge of within-category correlational structure but prior classification training did not (Chin-Parker & Ross, 2002; Sakamoto & Love, 2010). A limited awareness of within-category structure was reported by Little and Lewandowsky (2009, Experiment 2) when feedback in classification learning was probabilistic, rather than deterministic, presumably because probabilistic feedback resulted in attentional weights that were distributed more evenly across dimensions. Murphy and Ross (2010) demonstrated that the correlational structure was used to guide property judgments of novel instances when the stimulus displays were observable.

A related issue is whether the category label functions as a special feature. Yamauchi and Markman (2000) found across four experiments that subjects were more likely to endorse a feature value consistent with the category label when given an inference test. When given a classification test—that is, the full set of feature values were provided, and the category label was absent—the proportion of category accordance responses was determined by the number of feature matches with a prototype.

In the present study, category and feature inference and whether the category label functions as a special feature were addressed, but within the context of other variables that should be important: (1) whether inference is affected by the correlational structure of the critical dimensions defining each stimulus; (2) whether inference is further modulated by degree of category overlap (discriminability); and (3) whether the number of cues available at the time of test affects feature inference and classification in the same way. By dimensional structure, we mean whether the dimensions take on values that are highly correlated or uncorrelated with each other. By categorical structure, we mean whether the categories are separated or overlap with each other. Critically, the correlation between the dimensional values and the category label is larger when the categories have few overlapping features. By number of cues, we mean whether a later transfer test is composed of a stimulus set lacking one, two, or three features, where one of those features could be the category label itself. The theoretical importance of each of these manipulations is addressed in turn.

---

[1] This assumes that a stimulus can be defined by listing its presumed features, a description that is likely adequate when feature values vary along well-defined dimensions, as was used in the present study as well. However, this representation is likely inadquate when ill-defined categories (Neisser, 1967) are used in which the dimensions may be obscure and the appearance of an object may be driven by generative principles that are hidden (e.g., Gelman, 2004).

Logically, feature inference should be contingent upon the level of correlation among the dimensions defining each stimulus. Biological categories (and perhaps most others as well) likely have this property. Thus, nonvenomous snakes usually have a round pupil in the eye, whereas venomous snakes have a heavy triangular head and elliptical eyes; signs of diabetes include unexplained weight loss, unusual fatigue, and a tingling in the hands or feet; meat is likely spoiled if it smells rancid and has a slimy texture to it. That these features tend to co-occur is one reason why feature inference is possible; without these correlations, inference from a given set of features to an omitted attribute cannot occur. This assertion is assumed by Yamauchi and Markman (1998): "In inference, subjects tend to pay particular attention to relationships between exemplars within a category" (p. 125); and in their later study, Yamauchi and Markman (2000) stated that "current research identifies at least three crucial factors that govern inductive judgments using categories . . . (including) . . . correlation of features across category members" (p. 776).

Interestingly, the dimensions defining the categories in the experiments of Yamauchi and Markman (1998, 2000) were weakly correlated. In Yamauchi and Markman (1998), subjects observed four stimuli in each of two categories, with each stimulus defined by binary values along four dimensions; in Yamauchi and Markman (2000), there were five stimuli in each of two categories, with each stimulus defined by five binary dimensions. Table 1 summarizes these correlations, as well as specifying the various characteristics of each experiment. Although category 1 was primarily composed, schematically, of feature values 1 and category 2 by feature values 0, the dimensions themselves were minimally correlated.[2] However, each dimension was positively correlated with its category label. What seems likely is that the results of Yamauchi and Markman (1998, 2000) were, therefore, driven by the relationship between individual dimensions and the category label and not the relationship among dimensions.

In the present study, the correlations among the features on each dimension, as well as the correlation between the category label and any dimension, were manipulated. The reason it is critical to separate dimensional feature correlations from the correlation of the category label with its

---

[2] When variables are binary, the phi coefficient was computed, although the identical value is returned when computed as a standard Pearson product–moment correlation. These correlations are unaffected by the small number of patterns per category, since increasing the number of patterns per category (as in multiple learning trials) leaves the correlation unchanged as long as the same stimulus values are maintained. Similarly, substituting any continuously variable set of values around 1 or 0 minimally alters the dimensions to dimension correlation. In sum, the correlation is not the result of the use of a small number of binary valued stimuli; rather, it arises from the choice of structure.

Table 1 Correlations between category dimensions and between dimension and category label in Yamauchi and Markman (1998, 2000) and the present study

	Dim (x,y)	Dim x, Cat y
Yamauchi & Markman (1998)	$r = .333$	$r = .500$
Yamauchi & Markman (2000)	$r = .250$	$r = .600$
Present Study:		
Low OL, High Corr	$r = .729$	$r = .707$
Low OL, Low Corr	$r = .000$	$r = .707$
High OL, High Corr	$r = .729$	$r = .447$
High OL, Low Corr	$r = .000$	$r = .447$

Dim (x,y) is the correlation between dimensions of the same category; Dim x, Cat y is the correlation between a dimension and its category label. Dimension values were labeled 1–6, and category labels were labeled as 1 or 2. OL = category overlap

dimensions is that, at a minimum, feature inference should be driven by the former and label inference should be influenced by the latter. Figure 1 shows a schematic for a category representation. The instances of a particular category A experienced by the subject are represented by {A1, A2, . . . , A_n}; the features of these stimuli are represented by {f_1, f_2, . . . , f_n}. In theory, the features may be either correlated or not with the category label (r_{Af1}, r_{Af2} . . .) and may or may not be correlated with each other (r_{f1f2}, r_{f1f3}, . . .).

Figure 2 shows a schematic of the four between-subjects conditions in the present study, determined by dimensional correlations that were either high or zero and high or low category overlap. Although this figure captures the variables of dimensional correlation and category overlap, it fails to capture the fact that four, rather than two, dimensions defined each stimulus in the present study. Critically, when category overlap was minimal, the two categories were better separated, and the correlation between each dimension and its category label was high; when category overlap was moderate, this correlation was low. To calculate the correlations between dimensions and between category
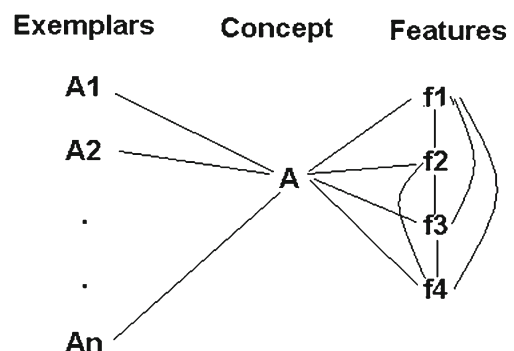

Fig. 1 Schematic representation of a category (A), the instances of category A in learning (A1, A2, . . .), and the features (f_1, f_2, . . .) that occur within category A
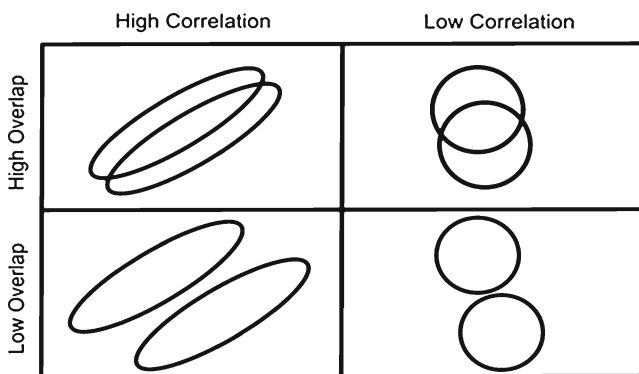
Fig. 2 Schematic representation showing categories with high or low dimensional correlation for categories that have low or high overlap. Both dimensions were manipulated in Experiment 1 and Experiment 2

labels and dimensions, each dimensional value was assigned a numerical label from 1 to 6, and category labels were assigned a numerical label, 1 or 2 (see the Appendix for values of learning stimuli). For the high dimensional correlation conditions, the correlation between any two of the four dimensions within each category varied from $r = .625$ to .875, with a mean correlation of $r = .729$; for the low-correlation conditions, these values ranged from $r = −.250$ to .250, with a mean correlation of $r = .000$. When the two categories were combined, the mean correlation between the two dimensions for the high-correlation condition was maintained ($r = .783$); for the uncorrelated, the overall correlation remained at zero. Importantly, these correlations were generated even though the same stimulus values were used in the high and low dimensional correlational conditions; only their ordering for each stimulus was varied to produce the requisite correlations.

In the low-overlap condition, the two categories were composed of dimensional values that were minimally shared; the resulting correlation between the category label and any dimension was moderately high, $r = .707$. In the high-overlap condition, the two categories were composed of dimensional values that were partially shared, resulting in a correlation between any dimension and its category label of $r = .447$. Manipulation of category overlap was achieved by adjusting the stimulus values on each dimension of one of the categories. Therefore, to produce the high-overlap categories, each dimension of one of the categories in the low-overlap condition was incremented by one value, thereby reducing its separation from the other category. The end result was that within-category similarity remained the same but between-category similarity was modified, thereby reducing the ratio of within-category to between-category similarity, a ratio that we have previously used to define category structure (Homa, Rhodes, & Chambliss, 1979). Importantly, the effect of reducing category overlap had the effect of increasing the

correlation between the category label and its dimensions.[3] We should note that the categories in the low-overlap condition were not linearly separable in any two dimensions but were in four dimensions. When the categories had high overlap, 2 of the 16 learning stimuli were slightly more similar to the alternate category.

Our initial hypothesis was that these two correlations—within-category dimensional correlation and the correlation of the category label with its dimensions—should selectively influence feature and category inference, respectively. In general, feature inference should be driven by dimensional correlation and category inference by overlap. The major caveat is that the two influences could interact with the number of cues available at the time of test. When the dimensions were correlated, feature inference was expected to be less reliant on number of features available at the time of test, since each feature is a moderate predictor of other feature values. When the dimensions were uncorrelated and the label was tested, we expected that number of features provided at the time of test would be more important for either or both of two reasons: With the number of cues increased, the test stimulus increasingly matches an item in memory; and the informational weight of the sum of independent cues should carry equal or more information than the sum of correlated cues.

In addition, the impact of dimensional correlation might be reduced if the category overlap was extreme at either end; if overlap is high, learning might be so difficult as to preclude better than chance performance on either a feature or a label test. In contrast, making the two categories so distinct would trivialize transfer, making all judgments too easy. Our goal was to gain empirical knowledge with this manipulation using a best guess of what overlap might be illuminating. F\inally, in each experiment, a brief recognition test of old, new, and prototype stimuli was provided following classification judgments on the transfer test. In

---

[3] The relationships between category overlap and the correlation between the dimensions of a category to its label are intrinsically bound together. Our previous measure of structural ratio of a category (Homa et al., 1979)—the ratio of within- to between-category distances or dissimilarities—is similar but not identical to the correlation of a category's dimensions to its label. If similarity is analytically defined by the number of feature matches between any two patterns, then the structural ratios for correlated low overlap, correlatead high overlap, uncorrelated low overlap, and uncorrelated high overlap were .447, .773, .475, and .868, respectively. When categories were defined as low overlap, the categories overlapped on each dimension but were linearly separable in four dimensions. When categories were defined as high overlap, the categories were not linearly separable, since two patterns in each category were closer to members of the alternate category. As was noted by one of the reviewers, the high- and low-overlap conditions share similarities to the probabilistic and deterministic feedback manipulation in Little and Lewandowsky (2009). As such, these authors would likely predict that within-category correlational structure should be realized to a greater extent under the high-overlap condition.

most inference studies, recognition judgments have not been used. Since a number of researchers (e.g., Chin-Parker & Ross, 2002; Murphy & Wisniewski, 1989) have argued that classification learning fosters awareness of categorically distinctive properties but not within-category correlational structure, recognition discrimination of old training patterns from novel patterns belonging to the same category should be poor or absent. Recognition accuracy, therefore, provides converging support that classification learning results in the storage of within-category structure above and beyond categorically distinctive properties, in the form of specific members stored in memory, a summary representation like a prototype, or sufficient partial exemplar knowledge that preserves some of the dimensional correlations.

Two experiments are reported that differed in the number of learning blocks prior to transfer. In Experiment 1, simultaneous arrays of all category members were presented for study, mirroring the procedure of Yamauchi and Markman (1998). In Experiment 2, the number of learning blocks was increased from 4 to 12 blocks, since we anticipated that the increased learning might foster transfer performance that reflected more strongly the influence of dimensional correlation and category overlap. Each experiment used stimuli that captured the morphological properties of bacteria, with each bacterium containing a membrane, a polar flagellum, nucleoid, and pili. Each dimension had six variations—for example, six increasing levels of membrane thickness. Figure 3 shows a typical stimulus and test trial in the learning phase. In each experiment, the subject was told to learn which characteristics defined the two classes of bacteria.

Experiment 1

All subjects received a booklet that contained the 8 stimuli belonging to category A on the left page and the 8 stimuli of B on the right side. After a study phase, the subject turned to the next page, which asked the subject to identify the category of each stimulus. There were four study/test blocks prior to transfer. Following learning, subjects received a transfer test requiring feature and label identification for 24 stimuli, followed by a brief recognition test.

Method

Subjects Two hundred thirty-four undergraduates at Arizona State University selected from introduction to psychology classes participated in the experiment. For the correlated and high-overlap, correlated and low-overlap, uncorrelated and high-overlap, and uncorrelated and low-overlap conditions, there were 55, 61, 61, and 57 subjects, respectively.
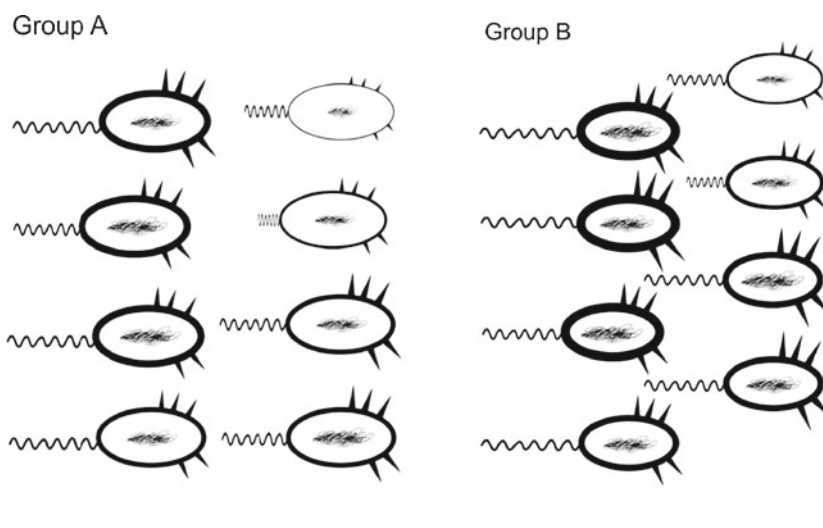
Materials and stimulus design Each bacterium contained four features: a membrane, polar flagellum, nucleoid, and pili. Six size variations of each feature were constructed. The incremental differences between the values were scaled using a variation of Weber's law where a value of 6 = 100 %, 5 = 90 %, 4 = 81 %, and so forth. The scaling was important to ensure that each value could be visually discerned as unique. All stimuli were created with iDraw to conform to exact pixel specifications. The pili of each bacterium were placed at five locations about the membrane. Bacteria were created using values for each of the four features, with a value of 1 being the smallest and 6 being the largest. Size values were selected to achieve either correlated or uncorrelated and high-overlap or low-overlap feature attributes. Each condition contained two categories of bacteria, those belonging to Group A and those belonging to Group B. A complete listing of the learning stimuli for the conditions of dimensional correlation and category overlap is contained in the Appendix.

Each booklet included instructions, four learning study–test trials, and the transfer test. In both learning and transfer, the subject selected either of two alternatives, either A or B in the learning phase, one of two numbers during the transfer test, and either old or new for the recognition test. During the study phase for each of the four learning trials, the stimuli for Groups A and B were arranged such that the two groups were side by side in the test booklet—each group entirely on its own page. The order of all stimuli was randomized for each learning study–test phase and transfer test for each condition so as to minimize order effects. Subjects were sequenced through the booklet at the same rate via navigation cues at the bottom right of each page (e.g., "continue" or "pause here").

Procedure Groups of up to 12 subjects were run simultaneously in a classroom setting. Subjects were allowed 1.5 min to study the eight bacteria belonging to each group (A and B) during the learning phase, followed by the test phase. In the test phase, subjects identified the same bacteria as belonging to Group A or B and recorded this information in their data booklet. This sequence was repeated three more times, for a total of four study–test phases.

Subjects were given 8 min to complete the 24 questions in the transfer portion of the experiment. During the transfer phase, subjects answered either feature inference or category inference questions. Counting the category label, each bacterium had five total features (membrane, flagellum, nucleoid, pili, and category label). During the transfer phase, bacteria missing one, two, or three features were presented. On feature inference questions, the category label feature was always present, and subjects had to choose one of two highlighted feature values for one of the missing features. For category inference questions, subjects responded with the appropriate category label value on the basis of the remaining features

Group A   Group B

_____

Sample Test Page Following Learning

Group A or B

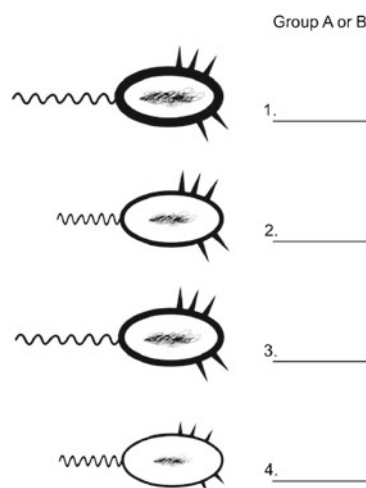

1._____

2._____

3._____

4._____

presented. On the feature test, one feature had a value that clearly placed it within the domain of its category; the other feature was always outside this domain. An example of a typical feature test is shown in Fig. 4, in which the category label and the tested feature are shown.

A brief recognition test was provided following the transfer classification test. On the recognition test, six stimuli were shown, three to a page, and the subject was instructed to simply indicate whether the stimulus had appeared in the learning phase (old) or not (new). Of the six stimuli, two were old (training), two were new, and two were the category prototypes. The presentation order of the stimuli on the recognition test was randomized.[4]

_____

[4] The transfer test also included confidence values and a final prototype test in which the subject was required to a select a best example for each category, one the modal prototype and one a near foil. However, the results of these tests were not informative across conditions. Consequently, these results are not reported.

Results

Learning Figure 5 shows the mean proportions of correct classification across the four learning blocks, separately for each condition. Each of the main effects was significant: Performance improved across blocks, $F(3, 690) = 30.32$, $\eta^2 = .116$, $MSE = 2.33$, and learning was affected by dimensional correlation, $F(1, 230) = 98.63$, $\eta^2 = .300$, $MSE = 4.10$, and category overlap, $F(1, 230) = 405.78$, $\eta^2 = .638$, all $ps < .001$. In addition, the correlational structure × level of overlap interaction was significant, $F(1, 230) = 43.30$, $\eta^2 = .158$, $MSE = 4.10$, $p < .001$, and reflected the fact that learning difference between high and low correlational dimensional structures was more affected by low-overlap structure than by high overlap. Although the level of improvement across blocks was slight, averaging about 8 %–10 %, terminal level of learning was moderately high, ranging from 70 % to 90 %. A Bonferroni subsequent test revealed that overall, learning accuracy was
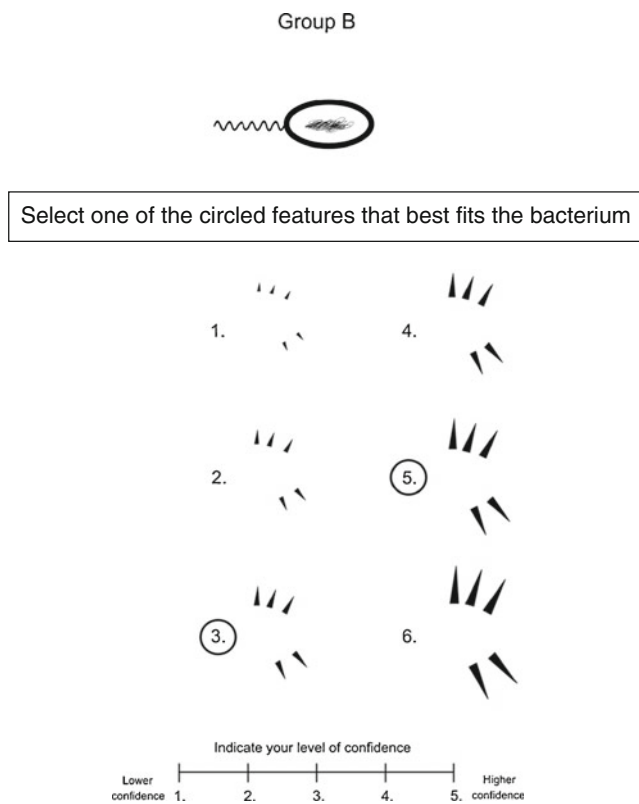
Group B

Fig. 4 Sample test following the learning phase, Experiment 1

ordered as follows: uncorrelated low overlap > correlated low overlap > uncorrelated high overlap = correlated high overlap ($p < .05$).

Transfer: classification Overall, accuracy on the transfer test was affected by condition, $F(3, 230) = 45.68$, $\eta^2 = .373$, $MSE = 3.769$, $p < .001$. A Bonferroni test with a
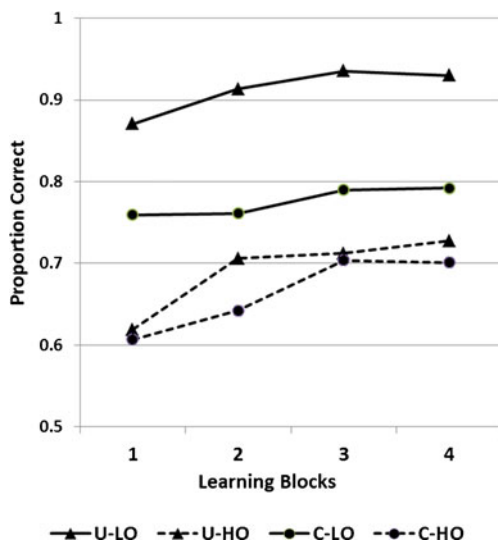

Fig. 5 Mean learning rate when categories contained correlated (C) or uncorrelated (U) dimensions for categories having low (LO) or high (HO) overlap, Experiment 1

significance level of .05 revealed that performance was ordered as follows: correlated low overlap (.868) > correlated high overlap (.818) = uncorrelated low overlap (.812) > uncorrelated high overlap (.641).

Structure, feature tested, and number of available cues on inference Figure 6 shows how overlap, dimensional correlation, and number of features available at test impacted inference; the left panel shows performance for the label test, and the right panel shows performance for the feature test. Overall, subjects were more accurate in inferring the category label (.808) than a category feature (.761), $F(1, 230) = 18.04$, $MSE = 2.05$, $\eta^2 = .073$, $p < .001$.[5] In addition, category overlap, $F(1, 230) = 55.41$, $\eta^2 = .199$, $p < .001$, and dimensional correlation, $F(1, 230) = 60.03$, $\eta^2 = .207$, were each significant, as was their interaction, $F(1, 230) = 16.44$, $\eta^2 = .067$, $MSE = 1.256$, all ps $< .001$. Overall, inference decreased by 10 % when category overlap was high and by 14 % when the dimensions were uncorrelated. The interaction between category overlap and dimensional correlation revealed that overlap had little impact on accuracy when the dimensions were highly correlated but significantly impacted transfer when the dimensions were uncorrelated. The effect of reducing the number of features at the time of test decreased performance by nearly 20 %, $F(2, 460) = 142.60$, $\eta^2 = .383$, $MSE = 0.48$, $p < .001$.

A major concern was whether the number of available features at the time of test differentially affected feature and label inference. The number of missing features interacted with both dimensional correlation, $F(2, 460) = 32.41$, $\eta^2 = .123$, $MSE = 0.48$, $p < .01$, and category overlap, $F(2, 460) = 7.35$, $\eta^2 = .031$, $MSE = 0.48$, $p < .01$. The most striking result was that the number of available features at the time of test affected performance more when the dimensional correlation was low. For the label test, performance on the high dimensional correlation conditions dropped by about 10 % when the number of missing features at the time of test was increased from one to three cues; when the dimensional structure was uncorrelated, performance dropped by about 30 %. A similar outcome occurred when inference of a feature was assessed: Performance dropped by about 10 % when the stimulus dimensions were correlated; when they were uncorrelated, decreasing the number of cues at the time of test decreased performance by about 25 %.

The impact of overlap was less striking. When the category label or category feature was tested, performance dropped by 24 % and 17 %, respectively, when the number of missing features increased from one to three and the category overlap was low; when the category overlap was

---

[5] When restricted to one missing feature, a condition that mirrored that in Yamauchi and Markman (1998), inference of the category label exceeded that of a category feature ($p < .01$).

Fig. 6 Mean accuracy on label inference (left panel) and feature inference (right panel), as a function of overlap, dimensional correlation, and number of features missing at test, Experiment 1
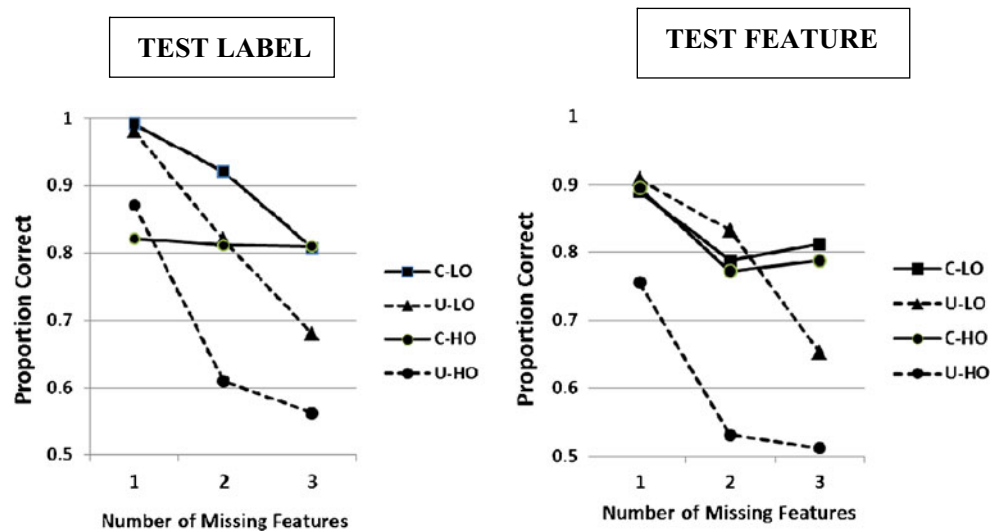
high, inference of the category label and category feature dropped by 16 % and 18 %, respectively, as the number of cues was reduced.

Recognition Regardless of condition, subjects discriminated old from new patterns with moderate accuracy; the proportion of old, new, and prototype stimuli called "old" was .725, .539, and .834, respectively. Table 2 shows the mean hit and false alarm rates for the old, new, and prototype stimuli as a function of training condition. Overall, recognition was more accurate following training on correlated dimension categories (mean hit rate = .778; mean false alarm rate = .505) than on uncorrelated dimension categories (mean hit = .672; mean false alarm = .572).

Discussion

Experiment 1 demonstrated that category and feature inference are influenced by both category overlap and dimensional structure. Additionally, each variable strongly interacted with the number of cues available at the time of test, such that cue

restriction had less impact when the category dimensions were correlated. In general, subjects were more accurate in inferring the category label than a feature, although the difference was slight. As was predicted, the number of available cues at the time of test had a reduced impact on accurate inference when the dimensions were highly correlated; this was true for the both label test and the feature test. When the category dimensions were composed of features that were uncorrelated, inference dropped substantially when the number of missing feature was increased. Category overlap, which modulated the correlation between the category dimensions and the category label, also impacted inference, being additive with dimensional correlation when the label was tested. When a feature was tested, category overlap interacted with dimensional correlation such that category overlap had little effect on feature inference when the dimensions were highly correlated. It was only when the categories were composed of uncorrelated dimensions that category overlap affected feature inference.

On the recognition test, subjects discriminated old from new stimuli, although this discrimination was reduced by dimensional correlation. For all conditions, the category prototype was identified as old more frequently than either the old or new stimuli. The fact that recognition discrimination between old and new stimuli was more accurate when the stimulus dimensions were correlated lends support to the hypothesis not only that category training sensitized subjects to the within-category structure, but also that this outcome was not due simply to memorization of the patterns; otherwise, subjects would have recognized patterns in the uncorrelated conditions as well as in the correlated conditions, which was not the case.

Before we discuss feature and category inference in more detail, Experiment 1 was replicated but with additional learning trials prior to transfer. One anticipated result that was not obtained was the importance of dimensional correlation on feature and label inference. In particular, we had anticipated that dimensional correlation would affect feature

Table 2 Probability that a transfer item was called "old," as a function of stimulus type, condition, and experiment

Condition	Experiment 1			Experiment 2		
	Old	New	Proto	Old	New	Proto
C-LO	.755	.490	.850	.725	.600	.875
C-HO	.800	.520	.850	.900	.525	.825
Correlated	.778	.505	.818	.812	.562	.850
U-LO	.695	.595	.840	.800	.625	.750
U-HO	.650	.550	.795	.675	.575	.825
Uncorrelated	.672	.572	.818	.738	.568	.838
Mean	.725	.539	.834	.775	.581	.819

inference more than it would affect inference of the category label. Although both factors produced strong main effects, the interaction between type of inference (label, feature) and correlational structure was not significant, $F < 1$. This outcome in Experiment 1 may have been due to the relatively restricted number of learning blocks prior to transfer. For subjects to acquire knowledge about the dimensional correlated structure of a category, additional learning blocks might be critical.

Experiment 2

Experiment 2 replicated Experiment 1, but with the number of learning blocks increased from 4 to 12 prior to transfer. In addition, subjects were run individually, and all stimuli were shown on a computer screen, rather than in booklets. Otherwise, the learning procedure and transfer test were identical to those used in Experiment 1.

Method

Subjects The subjects were 80 undergraduates drawn from the introductory psychology pool at Arizona State University, randomly assigned to one of four conditions: correlated high overlap, correlated low overlap, uncorrelated high overlap, and uncorrelated low overlap. The sole restriction was that 20 subjects were run in each condition, with approximately the same number of males and females in each condition.

Procedure The subjects either received 12 learning blocks or reached 100 % correct in any one block. Each learning block consisted of two parts, study and test. During the study part, subjects saw all 16 learning stimuli grouped by category. There were four different ordering of the stimuli, and each ordering was seen 3 times in random order. In each block, the subjects saw one ordering for 30 s. After the study part of each learning block, the learning stimuli were presented one at a time in a random order, and subjects had to classify the item as belonging to category A or B. Each response was followed by correct feedback for 1 s.

Immediately after learning, a two-part transfer test was given, with a format and procedure identical to those in Experiment 1: an initial inference test (12 with the label available and 12 with the label missing), followed by the recognition test. The 24 stimuli on the inference test were presented in a random order.

Results

Learning Figure 7 shows the mean proportion correct on each learning block, separately for each condition. As was the case in Experiment 1, both learning blocks and
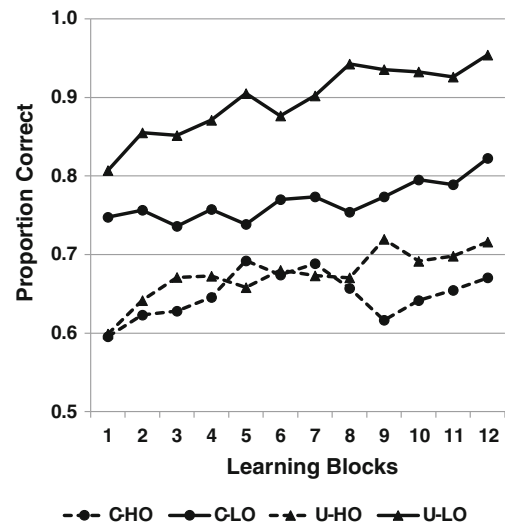

Fig. 7 Mean learning rate when categories contained correlated (C) or uncorrelated (U) dimensions for categories having low (LO) or high (HO) overlap, Experiment 2

conditions were significant (both ps $< .001$). A Bonferroni subsequent test ($p < .05$) revealed that learning accuracy was ordered as follows: uncorrelated low overlap > correlated low overlap > uncorrelated high overlap > correlated high overlap; this ordering mirrored that in Experiment 1, with terminal levels of learning ranging from .68 to .94. The terminal levels of learning due to the increased learning blocks were only marginally improved, relative to Experiment 1, and were confined to the low-overlap conditions.[6]

Transfer: classification Accuracy on the inference test for each condition mirrored that found in Experiment 1, with highest performance on the high correlation low overlap > high correlation low overlap > low correlation low overlap > low correlation high overlap (all ps $< .05$, Bonferroni). This ordering matched that in Experiment 1, with the exception that the high correlational structure condition with high overlap resulted in significantly better performance than did the low correlation. Overall, dimensional correlation enhanced inference, and high overlap reduced it (both ps $< .001$).

With few exceptions, the patterning of results replicated the findings of Experiment 1. Figure 8 shows mean inference accuracy as a function of type of test (label vs. feature),

---

[6] Only 3 subjects (out of 60) reached a learning criterion of 100 % in the correlated conditions and in the uncorrelated high-overlap condition. In contrast, nearly 50 % of the subjects achieved criterion in the uncorrelated low-overlap condition. However, number of correct classifications or recognition on the transfer test was uncorrelated with learning performance, where learning performance was defined by the overall accuracy of learning across the training blocks, either overall or assessed by condition. The unexpected finding that learning was easier in the uncorrelated conditions in each experiment, regardless of overlap, is addressed in the General discussion section.
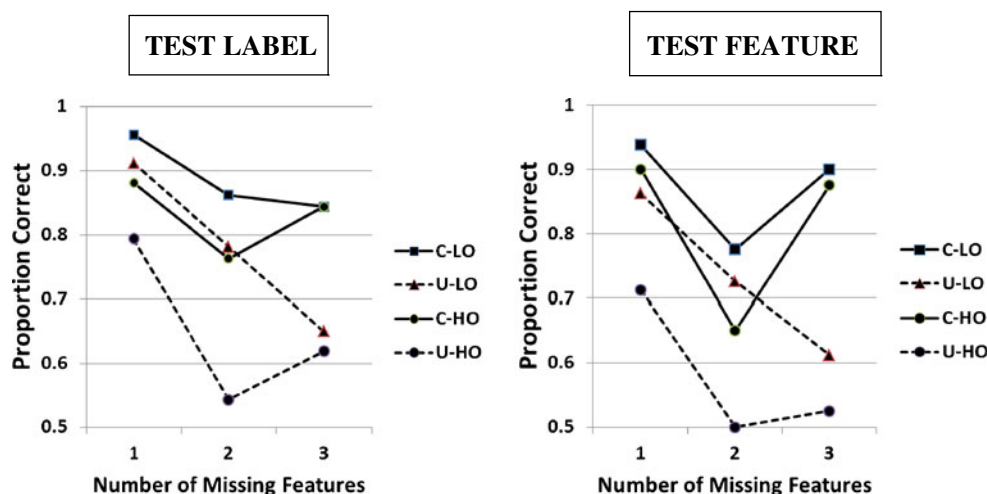
dimensional correlation, category overlap, and number of cues available at the time of test.

The main effect for each factor was statistically significant: number of features available at test, $F(2, 152) = 39.50$, $\eta^2 = .342$, $MSE = .029$; correlated versus uncorrelated dimensional structure, $F(1, 76) = 49.72$, $\eta^2 = .395$, $MSE = .049$; low versus high category overlap, $F(1, 76) = 21.94$, $\eta^2 = .224$, MSE .049, $p < .001$; and label versus no label, $F(1, 76) = 23.84$, $\eta^2 = .239$, $MSE = .032$, all $ps < .001$. The interaction between number of missing features and dimensional correlation was again significant, $F(2, 152) = 7.46$, $\eta^2 = .089$, $MSE = .029$, $p < .01$, as was the interaction between type of test (label vs. feature) and number of missing features at test, $F(2, 152) = 5.50$, $\eta^2 = .105$, $MSE = .030$, $p < .01$. When the features were uncorrelated, performance was increasingly degraded as the number of features available was reduced; when the category dimensions were highly correlated, the number of missing features at test had a substantially reduced impact on inference.

Experiment 2 produced two outcomes different from those of Experiment 1. First, the interaction between type of test (label, feature) and dimensional structure (correlated, uncorrelated) was significant, $F(1, 76) = 6.12$, $\eta^2 = .074$, $MSE = .032$, $p < .02$, reflecting the greater benefit of dimensional correlation on feature inference (.840 vs. 656) than on label inference (.879 vs. 777). Second, the interaction between dimensional correlation and overlap was not significant, $F(1, 76) = 2.88$, $MSE = .142$, $p = .09$. However, the patterning of performance was similar to that obtained in Experiment 1 in that overlap had a minimal effect on inference when the dimensions were correlated (.890 vs. .829) and a larger effect (.781 vs. .652) when they were uncorrelated.

Recognition The likelihood that the old, new, and prototype stimuli were called "old" on the transfer test is shown on the right panel of Table 2. As was the case in Experiment 1, subjects demonstrated reasonable accuracy in distinguishing old from new stimuli (.775 vs. .581), with the caveat that the

prototypes of each category were incorrectly called "old" at the highest rate (.819). Discrimination accuracy again favored the correlated conditions, as compared with categories composed of uncorrelated dimensions.

Discussion

The effect of increasing the number of learning blocks only slightly improved learning and was beneficial only when the category overlap was low. Nonetheless, Experiment 2 replicated the major findings of Experiment 1. In addition, the prediction that correlated dimensional structure would affect feature inference more than label inference was supported, perhaps reflecting the increased number of training blocks prior to transfer. Otherwise, the only other disparity was that the interaction between overlap and degree of dimensional correlation was significant in Experiment 1 but not in Experiment 2. However, even here, the patterning of performance was similar, with overlap having a reduced effect on inference when the dimensions were correlated and a larger effect when the dimensions were uncorrelated.

General discussion

Three major results were found in the present study. First, both dimensional correlation within a category and the degree of category overlap affected subsequent category and feature inference on a later transfer test. Although category overlap has been demonstrated to affect category classification (Goldman & Homa, 1977) and, possibly, the learning strategy adopted by subjects (Ell & Ashby, 2006), the role of dimensional correlation on feature and category inference is novel. In each experiment, transfer was most accurate overall when the category dimensions were highly correlated, and category overlap was low and poorest when the category dimensions were uncorrelated and category overlap was high. These

findings are at variance with numerous studies that have demonstrated sensitivity to within-category correlations following inference but not classification training (e.g., Chin-Parker & Ross, 2002; Sakamoto & Love, 2010). Second, restricting the number of cues available at the time of test for a category test or dimensional feature impacted performance, especially when the category dimensions were uncorrelated. This outcome is potentially tempered by the fact that the overall level of performance was quite high with minimal cues when the dimensions were correlated and, therefore, less opportunity for improvement existed with additional cues. Nonetheless, when the category dimensions were correlated, both classification and feature inference were high even with three missing cues; with additional cues, performance improved by an additional 10 %. In contrast, when the dimensions were uncorrelated, performance was near chance when the number of cues was minimal, especially when the categories overlapped to a moderate degree. Our initial expectation was that, in the uncorrelated dimension conditions, category identification, but not feature inference, would improve with additional cues, since a mild correlation existed between each feature and the label even in the high-overlap condition. We suspect that the reason that feature inference was increasingly enhanced with additional cues when the dimensions were uncorrelated with each other is due to the mediating role of the category label. That is, even with uncorrelated cues provided, each cue increasingly highlights the complex of cues associated with one category rather than the other; as category determination improves, so does inference of a missing feature, since these cues were associated to the category label during learning. Third, performance on identification of the category label was slightly but significantly better than identification of a category feature, regardless of category overlap, dimensional structure, or number of available cues at the time of transfer.

One reason why strong category and feature inference was revealed following category learning, whereas its effect was absent in numerous previous studies, may be because the category structure in our experiments contained patterns from correlated dimensions that were highly variable; that is, the values on each dimension were multivaried rather than binary. In previous studies exploring category and feature inference, small sets of binary-valued stimuli were used, and the correlation among the dimensions was minimal. For example, in the study by Sakamoto and Love (2010), subjects learned two categories either by inference or by classification training, followed by a transfer test. However, the computed correlation between the dimensions was weak; for example, the correlation between dimensions 1 and 2 was .403; between dimensions 2 and 3 and between dimensions 2 and 5, each correlation was .167. Although their results found greater sensitivity of the internal category structure following inference versus classification training, the internal structure in terms of correlated dimensions was weak. A similar analysis of other studies (e.g., Chin-Parker &

Ross, 2002; Yamauchi & Markman, 1998, 2000) reveals that the dimensions are, at best, weakly correlated with each other. Although inference training may better reveal the internal structure of categories, as compared with classification training, this conclusion may be restricted to small sets of binary-valued stimuli whose dimensions are weakly correlated.

We believe that the related conclusion, that classification training highlights only diagnostic category information and not its internal structure (e.g., Chin-Parker & Ross, 2002; Little & Lewandowsky, 2009), is also questionable. If only diagnostic information were stored following classification training, recognition of training patterns should be poor. In particular, subjects should be unable to discriminate old training patterns from new category patterns that contain identical diagnostic features. However, subjects in the present study made this discrimination with moderate accuracy, especially when the dimensions were highly correlated, regardless of the degree of category overlap. It is also likely that subjects stored more than the specific training stimuli, since recognition would then be driven more by category overlap than by dimensional correlation. As a result, recognition performance, combined with robust feature inference performance following classification, converges to the same conclusion: Classification training can foster knowledge of within-category structure. We should note that experiments using continuously variable stimuli, such as distorted forms (Homa, 1978), dot patterns (Posner & Keele, 1968, 1970), and schematic faces (Homa, Smith, Macak, Johovich, & Osorio, 2001), routinely find high levels of recognition following classification training (e.g., Homa, Goldhardt, Burruel-Homa, & Smith, 1993). It is also the case that the internal structure of categories is revealed by shaping variables that primarily highlight within-category structure. For example, when category size and pattern distortion are manipulated within the learning phase, transfer performance is dramatically changed (Homa, 1984). These variables are easily introduced with more variable stimuli; with binary-valued stimuli, shaping variables can, at best, be weakly incorporated into the experimental design. In conclusion, subjects in the present study clearly learned the internal, correlational structure of categories following classification learning, an outcome likely contingent upon the use of stimuli that are more complex than small sets of binary-valued stimuli.[7]

---

[7] An interesting exception was recently provided by Jones and Ross (2011), who used actual photographs of birds that belonged to one of six categories, where each category was associated with two features. Following classification or inference learning, subjects were required to classify photographs of novel members. Inference learners better classified photographs, suggestive of enhanced stimulus structure following infernce learning. However, accuracy in classification was poor in each condition (.26 vs. .19 in Experiment 1 and .28 vs. .19 in Experiment 2, where chance in each experiment was .167). These slight differences may have occurred because subjects in the inference condition viewed each photograph with its associated label for a longer period of time than did subjects in the classification condition.

The finding that the number of cues affected both category and feature inference was expected. However, the fact that category and feature inference were equally affected by this manipulation and that the number of cues impacted categories defined by uncorrelated dimensions was not. In general, when the dimensions were correlated with each other and with the category label, minimal cues were sufficient to accurately identify a queried feature or category label. In contrast, when the dimensions were uncorrelated with each other and weakly correlated with the category label, the reduction of cues produced performance that was near chance. The simplest explanation is that the subject acquires during learning both knowledge of cue values associated to each label as well as, when the dimensions are correlated, knowledge that each cue (feature) value is correlated with other features. When the category label is tested and the cues are restricted, evidence for a given category is low; with additonal cues, evidence is cumulated to improve the accuracy of judgment, modulated by the higher evidence when category overlap is low. When the dimensions are correlated and a feature test is required, subjects can use their knowledge that the feature value on a particular dimension or dimensions is associated the value of the queried feature. Again, the greater the number of cues, the more evidence that is available to make a judgment.

Less obvious is why increasing the number of cues was instrumental in improving feature inference when the dimensions were uncorrelated. One possible explanation is to assume that, during learning, the subject stores each category as a label that is embedded within a network of associated cues. At the time of test, this network of associations is activated either weakly or strongly on the basis of the number of provided cues. Even when the dimensions are uncorrelated, the category label is positively correlated with each dimension, especially when the category overlap is low. Identification of a feature then arises via mediation through the category label, and the more cues there are, the more the category label and its complex of associated dimensional values are activated

The issue of whether the category label is a special feature was one of the issues addressed by Yamauchi and Markman (2000). Evidence that the label functioned as a special feature was provided because classification—that is, identification of the category label—was more affected by a low similarity match in their study than was the inference of a category feature. However, two concerns should be noted. First, the subject made these decisions, not from memory and following a learning phase, but with a sample sheet available. That is, the subject was free to compare each stimulus with a visually available set of patterns belonging to the two categories. Whether the same response patterns would emerge following a learning phase involving these categories and testing that restricted decisions from memory was not addressed. Second, as was noted in the introduction,

an analysis of the categorical structure in their experiments revealed that the features were correlated to the category label, but not with each other. As a consequence, it is not surprising that the proportion of category accordance responses was different in the classification (category label) and feature inference conditions.

An alternative approach to this question can be addressed by two alternative criteria: Are subjects less likely to identify a feature than the category label, and does inference of the category label or feature interact with other variables that strongly determine inference? Given the confines of the present study, the answer is mixed. Regardless of condition—that is, whether the categories were composed of stimuli whose stimulus dimensions were correlated or not and whether the features were highly or weakly correlated with the category label—subjects significantly identified the category label with greater accuracy than a category feature. Although the difference favoring classification of the category label was slight in a numerical sense, this advantage was statistically significant in each experiment. In addition, the type of test (label vs. feature) interacted with stimulus structure and number of cues in each experiment. In effect, there was some evidence that the category label functioned unlike its features.

The theoretical importance of a category label might be better realized by using categories that include a generative component—that is, a prototypical form that constrains each of its members (Posner & Keele, 1968, 1970). Gelman has consistently demonstrated that young children ascribe to some categories a primitive and hidden essence that is distinct from its external features (Gelman, 2004; Rhodes & Gelman, 2009; Waxman & Gelman, 2009), although Deng and Sloutsky (2012) have shown that, for children and less so for adults, a particularly salient feature may guide inference. The assumption that categories possesss a hidden core or essence may arise implicitly when dimensions are correlated, perhaps further augmented when overlap is minimal. Consistent with this view is the finding by Billman and Knutson (1996), who found that inference was enhanced when the category was defined by three intercorrelated dimensions, as compared with a condition where the number of correlated dimensions was the same but were otherwise uncorrelated with each other. Their conclusion that complexity facilitates learning when the dimensions are organized in a coherent manner may arise because the intercorrelated complex generates a prototype-like representation or integrated unit that better maintains memory for later judgments that tap these relations. The introduction of shaping variables (Homa, 1984), such as category size, pattern variance, and so forth, and the manipulation of the types of associated features, such as causal links (Rehder & Kim, 2009), when combined with categorical structure variables as in the present study, might further elucidate conditions that foster the category label as a special feature. A theoretically productive line of research might be to build in feature

correlations, as in the present study, modified (perhaps by instructions) such that some, but not all, features were causal.

One unexpected finding was that categories composed of uncorrelated dimensions were learned slightly faster than those composed of correlated dimensions. This occurred regardless of overlap and in spite of clear evidence that inference on the transfer test was strongly facilitated by the correlational dimensional structure. Whether this reflects a general bias by subjects for unidimensional solutions (e.g., Ashby, Queller, & Berretty, 1999), which would preclude discovery of correlational structure, is unclear. Insight into this issue might be gained by systematically varying the degree of correlational structure in a categorical paradigm.

Finally, we agree with Yamauchi and Markman (1998) that a richer theory of categorization requires paradigms that ex-plore phenomena other than classification and that inference has received far too little focus. However, the claim that "distinct representations arise if people learn categories by inference or by classification" (p. 143) seems artificial, since it presumes that we can order our experiences accordingly. In effect, we accept whatever nature randomly provides—sometimes complete information requiring classification and, other times, partial information requiring feature inference. Given the chaotic and random temporal encountering of objects of all types in the world, positing two distinct representations, one marked as reflecting categorical judgments and the other as experiences based on feature inference, seems unlikely. Rather, we suspect that categorical knowledge arises as a primary outcome, with our representation supplemented by later judgments requiring feature inference.

Appendix

Table 3 Design of 64 stimuli used in all conditions

Condition	Stimulus number	Structure	Overlap	Group	Membrane	Flagellum	Nucleoid	Pili	Prototype
1	1	Correlated	High	A	1	2	1	1	3333
	2	Correlated	High	A	2	1	2	2	
	3	Correlated	High	A	3	3	4	4	
	4	Correlated	High	A	3	3	3	4	
	5	Correlated	High	A	3	4	3	3	
	6	Correlated	High	A	4	3	4	3	
	7	Correlated	High	A	4	4	3	4	
	8	Correlated	High	A	4	4	4	3	
	9	Correlated	High	B	2	3	2	2	4444
	10	Correlated	High	B	3	2	3	3	
	11	Correlated	High	B	4	4	5	5	
	12	Correlated	High	B	4	4	4	5	
	13	Correlated	High	B	4	5	4	4	
	14	Correlated	High	B	5	4	5	4	
	15	Correlated	High	B	5	5	4	5	
	16	Correlated	High	B	5	5	5	4	
2	1	Correlated	Low	A	1	2	1	1	3333
	2	Correlated	Low	A	2	1	2	2	
	3	Correlated	Low	A	3	3	4	4	
	4	Correlated	Low	A	3	3	3	4	
	5	Correlated	Low	A	3	4	3	3	
	6	Correlated	Low	A	4	3	4	3	
	7	Correlated	Low	A	4	4	3	4	
	8	Correlated	Low	A	4	4	4	3	
	9	Correlated	Low	B	3	4	3	3	5555
	10	Correlated	Low	B	4	3	4	4	
	11	Correlated	Low	B	5	5	6	6	
	12	Correlated	Low	B	5	5	5	6	
	13	Correlated	Low	B	5	6	5	5	
	14	Correlated	Low	B	6	5	6	5	
	15	Correlated	Low	B	6	6	5	6	

Table 3 (continued)

Condition	Stimulus number	Structure	Overlap	Group	Membrane	Flagellum	Nucleoid	Pili	Prototype
	16	Correlated	Low	B	6	6	6	5	
3	1	Uncorrelated	High	A	1	2	3	3	3333
	2	Uncorrelated	High	A	2	3	4	2	
	3	Uncorrelated	High	A	3	4	4	4	
	4	Uncorrelated	High	A	3	3	4	4	
	5	Uncorrelated	High	A	3	3	1	3	
	6	Uncorrelated	High	A	4	4	3	1	
	7	Uncorrelated	High	A	4	1	3	3	
	8	Uncorrelated	High	A	4	4	2	4	
	9	Uncorrelated	High	B	2	3	4	4	4444
	10	Uncorrelated	High	B	3	4	5	3	
	11	Uncorrelated	High	B	4	5	5	5	
	12	Uncorrelated	High	B	4	4	5	5	
	13	Uncorrelated	High	B	4	4	2	4	
	14	Uncorrelated	High	B	5	5	4	2	
	15	Uncorrelated	High	B	5	2	4	4	
	16	Uncorrelated	High	B	5	5	3	5	
4	1	Uncorrelated	Low	A	1	2	3	3	3333
	2	Uncorrelated	Low	A	2	3	4	2	
	3	Uncorrelated	Low	A	3	4	4	4	
	4	Uncorrelated	Low	A	3	3	4	4	
	5	Uncorrelated	Low	A	3	3	1	3	
	6	Uncorrelated	Low	A	4	4	3	1	
	7	Uncorrelated	Low	A	4	1	3	3	
	8	Uncorrelated	Low	A	4	4	2	4	
	9	Uncorrelated	Low	B	3	4	5	5	5555
	10	Uncorrelated	Low	B	4	5	6	4	
	11	Uncorrelated	Low	B	5	6	6	6	
	12	Uncorrelated	Low	B	5	5	6	6	
	13	Uncorrelated	Low	B	5	5	3	5	
	14	Uncorrelated	Low	B	6	6	5	3	
	15	Uncorrelated	Low	B	6	3	5	5	
	16	Uncorrelated	Low	B	6	6	4	6	

References

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61,* 1178–1199.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 458–475.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (Eds.). (1966). *A study of thinking*. New Brunswick, New Jersey: Transaction Publishers.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition, 30,* 353–362.

Deng, W., & Sloutsky, V. M. (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science, 23,* 178–186.

Ell, S. W., & Ashby, F. G. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception & Psychophysics, 68,* 1013–1026.

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences, 8,* 404–409.

Goldman, D., & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 375–385.

Homa, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 407–416.

Homa, D. (1984). On the nature of categories. *Psychology of Learning and Motivation, 18,* 49–91.

Homa, D., Goldhardt, B., Burruel-Homa, L., & Smith, C. (1993). Influence of manipulated category knowledge on prototype classification and recognition. *Memory & Cognition, 21,* 529–538.

Homa, D., Rhodes, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory, 5*(1), 11–23.

Homa, D., Smith, C., Macak, C., Johovich, J., & Osorio, D. (2001). Recognition of facial prototypes: The importance of categorical structure and degree of learning. *Journal of Memory and Language, 44,* 443–474.

Hull, C. I. (1920). Quantitative aspects of the evolution of concepts. *Psychological Monographs, 28* (Whole No. 123).

Jones, E. L., & Ross, B. H. (2011). Classification versus inference learning contrasted with real-world categories. *Memory & Cognition, 39,* 764–777.

Lassaline, M. E., & Murphy, G. L. (1996). Induction and catcategory coherence. *Psychonomic Bulletin & Review, 3,* 95–99.

Little, D. R., & Lewandowsky, S. (2009). Better learning with more error: Probabilistic feedback increases sensitivity to correlated cues in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1041–1061.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37–50.

Murphy, G. L., & Ross, B. H. (2010). Category vs. object knowledge in category-based induction. *Journal of Memory and Language, 63,* 1–17.

Murphy, G., & Wisniewski, E. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23–45). Chichester, England: EllisHorwood.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Rehder, B., & Kim, S. (2009). Classification as diagnostic reasoning. *Memory & Cognition, 37,* 715–729.

Rhodes, M., & Gelman, S. A. (2009). Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychological Bulletin and Review, 16,* 920–924.

Sakamoto, Y., & Love, B. C. (2010). Learning and retention through predictive inference and classification. *Journal of Experimental Psychology: Applied, 16,* 361–377.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences, 13,* 258–263.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124–148.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 776–795.