

# How goals affect the organization and use of domain knowledge

BENJAMIN D. JEE AND JENNIFER WILEY  
*University of Illinois, Chicago, Illinois*

Expert specialists organize their knowledge around information related to their goals. In the experiments presented here, the relation between goal use and knowledge organization was investigated by manipulating participants' goals as they learned about a novel domain. Experiment 1 showed that goal use produces biases toward goal-related information in categorization and induction. Experiment 2 revealed that the bias toward goal relatedness is not absolute; participants use their knowledge flexibly, depending on the context of induction. Experiment 3 showed that using information in the absence of a meaningful goal does not produce significant goal-related biases. Altogether, the effects of goal use are evident across a number of tasks and may depend on goal meaningfulness and the coherence it provides to goal-related knowledge structures.

Our experience of the world is mediated largely by the ways in which we categorize objects. What determines the categories that we learn and use? A great deal of research on human categorization has been done to examine how category structure affects this learning—for example, contrasting the ease of learning for disjunctive versus conjunctive categories, one- versus multidimensional categories, or linearly versus nonlinearly separable categories (e.g., Bruner, Goodnow, & Austin, 1956; Medin & Schwanenflugel, 1981; Shepard, Hovland, & Jenkins, 1961). Such research has produced a large body of evidence demonstrating that category structure plays an important role in category learning, and our understanding of this role becomes further refined with continued experimental efforts. In contrast to this predominant approach, some recent research has shifted the focus away from the structure of the to-be-learned categories and onto the characteristics of the learner. The present work extends this line of research by investigating the effects of the learner's goals on the acquisition and organization of domain knowledge. Here, goals are considered as some end-state toward which the learner is striving (cf. Newell & Simon, 1972)—for example, determining foods to be eaten on a diet or gifts appropriate for a 2-year-old. Across three experiments, we presented different groups of participants with different goals as they learned about the same set of novel stimuli. We then investigated the effects of these varying goals by administering subsequent categorization and induction tasks.

## Goals and the Organization of Domain Knowledge

There have been a number of studies demonstrating that expert individuals categorize domain-related information

differently than do nonexperts (e.g., Chi, Feltovich, & Glaser, 1981; Murphy & Wright, 1984; Tanaka & Taylor, 1991). The general picture that emerges from this research is that the transition from novice to expert involves the acquisition of progressively “deeper” knowledge structures (Bédard & Chi, 1992). However, when we examine differences *between* experts with different specializations in a domain (which are effectively overlooked in expert–novice comparisons), we find evidence that there may be more to this story.

Recent research has shown that experts organize their domain knowledge around ideals—that is, characteristics that best serve their goals (Burnett, Medin, Ross, & Blok, 2005; Lynch, Coley, & Medin, 2000; Medin, Lynch, Coley, & Atran, 1997; Shafto & Coley, 2003). Medin et al. (1997), for example, found that different types of tree experts (taxonomists, landscapers, and maintenance workers) categorized the same set of familiar tree species in distinctive ways. Not surprisingly, the taxonomists' categorizations aligned closely with scientific structure. The other groups of experts, however, appeared to form categories based on properties relevant to their particular goals in interacting with trees, such as weediness, landscape utility, and aesthetics. Similar variability between expert specialists has been documented with genetics experts (Smith, 1990, 1992) and experts in aquatic systems (Hmelo-Silver & Pfeffer, 2004). Experts may even rate the goodness of example (GOE) for exemplars in terms of similarity to ideals, rather than in terms of such factors as familiarity and central tendency (Burnett et al., 2005; Lynch et al., 2000). Thus, for some tree experts, the best examples of trees may have atypical attribute values, such as extreme height. Altogether, these findings suggest that goals have important consequences for expert knowledge,

---

B. D. Jee, bendj@uic.edu

with different specialists organizing their knowledge around different ideals.

The influence of goals does not appear to be restricted to highly trained experts; it may be a factor in the organization of common knowledge as well. Barsalou (1983) found that people utilize so-called goal-derived categories for planning and reasoning (e.g., things to take out of the house in the event of a fire). These goal-derived categories seem to violate the correlational structure of the environment, in that their members bear little resemblance to one another (e.g., the aforementioned category may include expensive paintings, children, pets, and family heirlooms). Nevertheless, people demonstrate an impressive consensus in their GOE ratings for members of goal-derived categories, and these ratings maintain stability over time (Barsalou, 1985). In Barsalou's (1985) research, participants appeared to base their GOE ratings on ideal feature values related to the goal(s) under consideration; the closer the exemplar's feature values were to these ideal values, the higher its *goodness* rating. These results mirror the findings with expert specialists, who appear to use ideals in their GOE ratings for domain-related stimuli.

Interestingly, Barsalou (1985) also found that ideals contributed to participants' GOE ratings for exemplars in common *taxonomic* categories, such as birds, furniture, and so forth. For example, the ideal of *how much people like it* accounted for a significant amount of unique variance in participants' GOE ratings for exemplars from the category of birds. Similarly, Ross and Murphy (1999) found that people categorized familiar food items not only in terms of familiar food group distinctions (e.g., meats, dairy products, etc.), but also by apparently goal-based standards that dictate when they are eaten (e.g., breakfast foods, snack foods, etc.). The picture seems to be that goals have far-reaching effects on knowledge organization; experts are not the exception.

Although highly suggestive of the influence of goals during learning, the findings considered thus far are all correlational in nature. Thus, they do not rule out the possibility that some factor(s) other than goal use could account for the observed variations in knowledge organization, both between expert specialists and within normal populations. Our experimental procedure was specifically designed to investigate the effects of goal use on knowledge acquisition, while controlling for other variables that might affect category learning, such as the frequency of exposure to stimuli, the nature of the learning task, and, of course, the structure of the presented stimuli. We also manipulated goals across experiments, thus enabling us to explore what properties of the goal were responsible for their influence on domain knowledge. Because numerous studies point to the influence of goals on knowledge acquisition, this experimental approach provides an important bridge to previous research.

### The Effects of Interacting With Stimuli

Exploring the influence of goal variation on categorization and induction also extends recent research that has examined how category knowledge is affected by different interactions with category members. One such line

of research has compared how learning differs between traditional category learning tasks and other tasks, such as inference tasks, which require the learner to infer an exemplar property on the basis of its category membership and other known properties (Yamauchi, Love, & Markman, 2002; Yamauchi & Markman, 1998). The findings from this research suggest that learners develop different knowledge depending on the task constraints that govern their interactions with exemplars. For example, Yamauchi and Markman found that learning through inference led to the development of category knowledge that emphasized the internal structures of the learned categories, whereas learning through classification led to knowledge emphasizing the features that distinguished between them.

Another line of research has examined how previously acquired category knowledge is affected when exemplars are subsequently used in some additional task (Ross, 1996, 1997, 1999, 2000). The findings from this research suggest that the learner establishes a single knowledge structure that is shaped by both classification and use, rather than forming separate representations for each task (Markman & Ross, 2003). For example, Ross (2000) found that the use of particular disease symptoms to prescribe medical treatments led participants to favor these features in a subsequent classification task, even though they were no more diagnostic of category membership than were other features that accompanied them.

Our experiments extended this research by examining the effects of interacting with stimuli in an *unsupervised* learning environment, in which the learner received no explicit instruction or feedback about the category structure of the stimuli. Most previous studies have used *supervised* learning tasks—for example, the standard concept attainment task (cf. Bruner et al., 1956), which typically involves presenting the learner with exemplars from different categories one by one, asking them to make a category decision for each exemplar, and then giving them feedback about the accuracy of each response. Thus, category information is often explicitly taught prior to the interaction manipulation. Rather than examining the effects of goal-directed interactions on preestablished category knowledge, we examined how interactions affect the emergence of different knowledge structures. Unsupervised learning is arguably a more ecologically valid approach by which to examine the effects of goals on knowledge organization (see Murphy, 2002, 2003), especially when one considers that expertise develops through years of experience in a domain, rather than through a process of classification followed by external feedback.

In summary, a number of correlational studies have suggested that goals influence how experts and nonexperts organize their domain knowledge, yet little experimental research has been done to investigate how goal use affects the emergence of goal-based knowledge structures. The purpose of the present research was to explore how goal-directed learning affects the acquisition and organization of domain knowledge. In addition to examining the effects on knowledge organization by using a categorization task, we also examined how participants applied their domain knowledge by presenting them with a property induction

task. If goal use results in category knowledge that is biased toward goal-relevant ideals, then, in certain contexts, property induction may evidence a similar bias.

### Overview of Experiments

In the three experiments presented here, participants were exposed to the same stimuli (a set of novel creatures) but interacted with them according to different goals. These interactions required the participants to determine which of the stimuli were best suited for attaining their designated goal. After a two-session phase of goal-directed interactions with the stimuli, the participants were given a hierarchical sorting task, followed by a property induction task. The entire procedure took four experimental sessions, because the participants were required to sort the stimuli both before and after the 2-day interaction phase, allowing us to assess changes in knowledge organization directly.

It was hypothesized that the nature of the participants' goal-directed activities would influence their categorizations and inductions. If the participants organized their knowledge of the stimuli around goal-related features, an idealized category structure based on such features should provide a good fit to the categories that they construct. If the participants utilized goal-related information for property induction, they should generalize a novel property from one item to another item that was similar with respect to goal-relevant features.

### EXPERIMENT 1

The purpose of the first experiment was to experimentally manipulate the participants' goals as they learned about a novel set of creatures and to assess the effects of this manipulation on their domain knowledge. All the participants were presented with the same cover story, that they had been hired as consultants for a corporation. The goal for this consultation position was manipulated between groups, but in all the conditions, the participants interacted with the same set of creatures. Each of these creatures had three types of properties in equal proportions. Two groups of participants were assigned to a goal condition that required them to use one of these feature types, the particular type differing between groups. A control group was assigned to a goal condition that did not require the use of any particular feature type over another. Thus, different feature types were goal relevant for the former two groups, whereas no particular type was goal relevant for the control group.

In the first experiment, the participants were presented with two postlearning tasks: a hierarchical categorization task and a blank property induction task. In the hierarchical categorization task, the participants were required to sort the stimuli into pairs, then groups of four, and then groups of eight. In the blank property induction task, the participants were required to generalize a novel property from one stimulus to one of two options. This novel property was a so-called *blank* property, designed so that it would be unfamiliar to the participants and could not be straightforwardly related to particular stimulus features

(see Osherson, Smith, Wilkie, López, & Shafir, 1990; Rips, 1975).

### Method

#### Participants

Twenty-one undergraduates from the University of Illinois at Chicago participated to fulfill a course requirement. The participants were run individually, and 7 participants were randomly assigned to each of three conditions.

#### Goal Manipulation

Goals were manipulated between groups by providing different versions of a cover story. All the participants were informed that a large corporation had purchased an island habitat with the intention of introducing a set of 16 creatures to it. The participants were told that the corporation had hired them as consultants. Specifically, the avoid predators group (Group AP) was told that the corporation hired them to determine which creatures were best suited for avoiding predators, the nutritional value group (Group NV) was told that the corporation hired them to determine which creatures were highest in nutritional value, and the control group (Group C) was told that the corporation was interested in monitoring the creatures and had hired them to determine which would be easiest to remember. The participants pursued their designated goals in the context of the interaction tasks described below.

#### Stimuli

On the basis of stimuli from a previous experiment (Jee & Wiley, 2004),<sup>1</sup> a set of 16 creatures was constructed using three feature types: pictorially represented (PR) features, features related to avoiding predators (AP), and features related to nutritional value (NV). PR features were depicted in an illustration of the creature, and AP and NV features were written below the illustration. (Figure 1 displays two sample creatures; Appendix A provides written descriptions of the PR properties and information about their distribution.) The AP and NV features were designed to have no prior meaningfulness to the participants (e.g., *izzdedd cells*). However, their meanings were embedded within each feature label; for example, possessing *izzdedd cells* allowed a creature to *play dead* (Appendix B lists each feature label and its meaning). The reason that the AP and NV feature labels were presented, instead of their meanings, was to eliminate the pos-

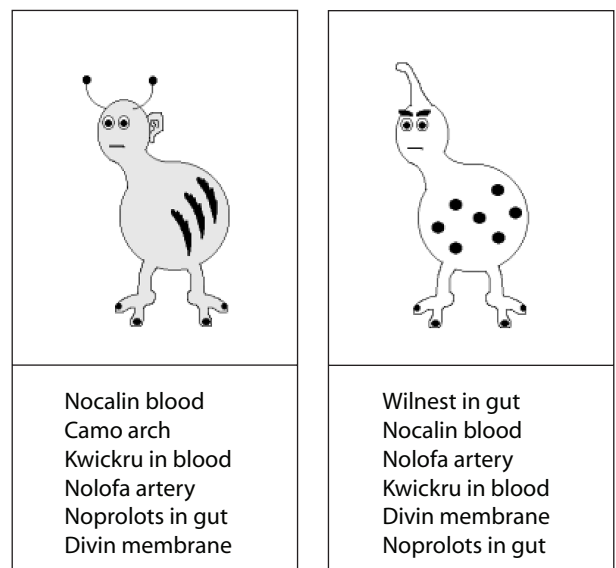


Figure 1. Two examples of the creature cards used in these experiments. The actual cards were in color.

sibility that the participants would use prior knowledge of these features in their initial categorizations; this would obscure our pre- to postlearning comparison and could interfere with their performance in the goal-directed interaction tasks. To avoid these unwanted possibilities, the feature labels were presented in lieu of their meanings throughout the experiment. The participants were taught all of the feature meanings in the pretraining tasks outlined below.

In order to test our hypotheses about the effects of goal-directed activity on knowledge organization, the PR, AP, and NV features were distributed among the creatures so that unique hierarchies would be formed using each feature type. Each creature had 9 features altogether, 3 of each type. These 3 features were selected from a pool of 14 features, which differed in frequency: 8 features occurred in 2 of the 16 creatures, 4 features occurred in 4 of the 16, and 2 features occurred in 8 of the 16 (the Appendixes display each feature, along with its frequency). Each creature possessed 1 feature of each frequency for each type (the stimuli possessing each property are indicated in the Appendixes). This distribution permitted the creatures to be sorted hierarchically, using each feature type by grouping them on the basis of feature commonalities at each frequency level: into eight pairs using the lowest frequency features, into four groups of 4 using the next lowest frequency features, and into two groups of 8 using the highest frequency features. The Appendixes show that different categories will be formed depending on which feature type is used as a basis for classification; for instance, Creature 1 is most similar to Creature 4 in terms of PR features, to Creature 2 in terms of AP features, and to Creature 3 in terms of NV features.

Another important characteristic of both the AP and the NV features is that half of them were positive (e.g., *camouflage capability*, *rich in vitamins*), and half were negative (*no camouflage capability*, *not rich in vitamins*). The negative features were designated by the prefix *no* in the feature labels (e.g., *Camo arch = camouflage capability*, *Nocamo arch = no camouflage capability*). Eight of the 16 creatures received positive AP features (Creatures 1–8), and 8 received negative AP features (Creatures 9–16). Likewise, half of the creatures received positive NV features (the odd-numbered creatures), and the other half received negative NV features (the even-numbered creatures). Consequently, half of the creatures with positive AP features had positive NV features (Creatures 1, 3, 5, and 7), whereas the other half had negative NV features (Creatures 2, 4, 6, and 8). Half of the creatures with negative AP features had positive NV features (Creatures 9, 11, 13, and 15), whereas the other half had negative NV features (Creatures 10, 12, 14, and 16). This distribution of positive and negative features implies that the sorting structures based on AP and NV features not only will be structurally unique, but also will contain distinct groups and subgroups for creatures with positive and negative features of each type.

### Overview of Procedure

As has been stated, goal information was varied between groups, but each group was exposed to the same stimuli. Before receiving goal information, the participants performed their first sorting of the stimuli (the *hierarchical sorting task*) and then learned the meanings of the feature labels displayed with each creature's pictorial representation, followed by a short test on this information (the *preinteraction tasks*). This all took place during the first experimental session. In the second and third sessions, the participants were required to use their goal in their interactions with the stimuli (the *interaction tasks*). The interaction tasks were preceded by a short test to assess whether the participants remembered the meanings of all of the feature labels that they had learned about in Session 1, not just the goal-relevant ones. In Session 4, the hierarchical sorting task was administered again, followed by the *blank property induction task*. We will describe each of the experimental tasks below.

**Preinteraction tasks.** The participants learned the meanings of the feature labels in a self-paced slide show. Following this, they completed a computerized multiple-choice test consisting of 14 questions about the feature meanings. Since the negative feature values simply contained the positive feature label preceded by

the prefix *no*, we considered it redundant to test the participants' memory for all 28 meanings. Each feature label was tested, either in positive or negative form. The participants were given up to five test attempts to complete the test without error. Three versions of the test were constructed by manipulating the order of the questions and the order of options within each question. A different test version was administered to each participant on Days 1, 2, and 3.

**Interaction tasks.** These tasks contained memory and evaluation components, in which the participants were expected to identify and use goal-relevant information. At the beginning of each trial, a creature (illustration and written features; see Figure 1) was displayed on a computer screen for 10 sec. Following this display, the participants completed a 2-page form. On page 1 was a checklist of all 28 possible creature features, and the participants were instructed to check the features possessed by the creature that had just been displayed. The participants in Groups AP and NV were asked to check only the features relevant to their respective goals, whereas the participants in Group C were asked to check all of the creature's features. The purpose of this memory component of the task was to determine whether the participants could select and correctly remember the goal-relevant features belonging to the creature that they would have to evaluate. On page 2 of the form, the participants evaluated the presented creature with respect to their goal: whether it would be good at avoiding predators (Group AP), a good source of nutrition (Group NV), or difficult to recall later on (Group C). After completing both pages, the participant advanced the screen, initiating the next trial. Each of the 16 creatures was shown in a random order.

**Hierarchical sorting task.** Each of the 16 creatures was displayed on a 5 × 8 in. card, including both its pictorial and its written features. Before sorting the cards, the participants were shown each of them one by one for about 2 sec apiece. Following this, the cards were randomly arrayed on a table in front of the participants, and they were instructed to find the "best way" to organize them. First, the participants were asked to form eight pairs of creatures. They were then asked to form four groups of 4 creatures, using the pairs that they constructed. Finally, they were asked to combine the four groups of creatures into two groups of 8. This task was administered on Day 1 and Day 4. The sorting instructions on Day 4 were identical to those given on Day 1, except that on Day 4, the participants were instructed to use their knowledge of the creatures to construct categories.

**Blank property induction task.** In each trial of this task, the participants were informed that a given creature had a novel property, "sarca." This property was selected to have no particular meaningfulness to the participants and to have no privileged relation to PR, AP, or NV features. The task was to select which of two target creatures was most likely to also have the novel blank property, given that the base creature had it. All the items were presented on a computer screen. The original item was presented for 10 sec and then replaced by two targets that were displayed until the participants made their selection. Following their response, a blank screen was displayed for 2 sec, followed by the next trial.

The participants received two types of trials. In both trial types, a more pictorially similar target (the PR target) competed with either a more AP-related or a more NV-related target. In AP versus PR trials, one target shared more PR features with the original item, and the other target had more AP features in common with the original. Importantly, both targets had exactly the same NV-related features in common with the original. Thus, these trials tested whether induction of the novel property would be based on pictorial or AP-related commonalities between the original item and the targets. An example of an AP versus PR trial could have Creature 2 as the base and Creatures 8 and 10 as the targets. In terms of AP features, Creature 2 is more similar to Creature 8, because both have three positive AP properties and share the ability to dive underwater. In terms of PR features, Creature 2 is more similar to Creature 10, because both have stripes on their bodies. In terms of NV features, however, Creature 2 is equally similar to Creatures 8 and 10, because all three creatures have three negative NV features and do not have a lot of protein. The NV versus PR trials were analogously designed so that

a target with more NV features in common with the base would compete against one with more PR commonalities, while controlling for AP similarities. The participants were presented with an equal number of AP versus PR and NV versus PR trials. Trial order was randomized to create a single task list of 34 trials.

### Procedure

The experiment took place over 4 consecutive weekdays. In Session 1, the participants performed the hierarchical sorting task for the first time (as a prelearning baseline), completed the preinteraction tasks (learning the feature labels and completing a test on this information), and then received goal information specific to their group assignment. Sessions 2 and 3 had formats that were similar to each other, with the participants first completing a test on the meanings of the feature labels and then completing the interaction task. Different versions of the feature test and interaction task were presented on each day. Finally, in Session 4, the participants completed the hierarchical sorting task for the second time (with the instruction to use their knowledge of the creatures to sort them) and then performed the blank property induction task. Each session took under 1 h to complete.

## Results

### Effects of Goal Use on Categorization: Hierarchical Sorting Task

The participants were required to sort the stimuli twice; once before the goal-directed interaction phase to establish a baseline and once afterward to examine the effects of these goal-directed interactions. Treatment of the sorting data closely followed techniques used by Medin et al. (1997). Each sorting hierarchy was used to derive a  $16 \times 16$  pairwise distance matrix, yielding an initial and final distance matrix for each participant. A number was assigned to each cell of the matrix corresponding to the level at which the pair of creatures was combined. Creatures combined at the lowest level (when the participant formed pairs) were assigned a distance of 1; creatures combined at the next level (groups of four) were assigned a distance of 2; creatures combined at the highest level (groups of eight) were assigned a 3; creatures *not* combined at any level were assigned a default distance of 4. Only the 120 unique cells above the diagonal were used in subsequent analyses, because of their redundancy with the cells below and because the cells on the diagonal (the distance between a creature and itself) are irrelevant. Distances were averaged across participants in each group to create six different matrices, representing each group's initial (Day 1) and final (Day 4) sorting. Distance matrices were also derived from the ideal sorting structures (APi and NVi), as well as the ideal based on PR features (PRi). The initial and final distance matrices for each group were correlated with the matrices derived from each of the ideal structures. Before presenting the correlations between the group and the ideal matrices, it must be noted that as a result of the distribution of each feature type in the stimulus set, the ideal structures were not significantly correlated with one another (these correlations ranged between  $-.16$  and  $.15$ , *n.s.*). This implies that a high correlation between a group structure and an ideal structure cannot be attributed to covariance among the ideal structures.

Table 1 displays the correlations between each group's initial and final distance matrices and each of the ideal matrices. (Individual participants' data will be displayed later,

in the Results section for the second experiment.) A similar pattern of correlations is revealed in the initial sorts for all three groups, with pictorial commonalities being highly favored. This initial pictorial bias is no longer evident in the participants' final sorts, although Group C's final correlation with PRi approximates their correlations with the other two ideals, APi and NVi. The pattern of correlations for the final sorts also reveals important differences between groups. Specifically, Groups AP and NV formed category structures that closely aligned with their respective ideals, APi and NVi, whereas Group C showed a modest correlation with each structure. This pattern of results was further explored by determining which ideal structure, APi or NVi, provided a better fit to each group's final sort.

Following procedures outlined by Pellegrino and Hubert (1982), these analyses of fit involved normalizing the APi and NVi distance matrices, using these normalized matrices to compute a difference matrix, and then correlating this difference matrix with each group's final distance matrices. A significant correlation between the difference matrix and a group matrix would imply that one of the ideal structures provided a better fit to the data than did the other structure. These analyses of fit revealed that APi provided a significantly better fit than did NVi for Group AP's final sort [ $r(120) = .61, p < .01$ ], NVi provided a better fit for Group NV's final sort [ $r(120) = .59, p < .01$ ], and APi and NVi fit Group C's final sort equally well [ $r(120) = .10, n.s.$ ].

In summary, all three groups evidenced the initial tendency to use the pictorially represented features for sorting the creatures on Day 1, as could be expected; at this point, the written features were nonsensical. However, the sorts on Day 4 evidenced the influence of goal-directed activity on category formation. All the groups shifted away from using the PR features in their final sorts, but only Groups AP and NV showed a strong shift toward another ideal, the ideal related to their respective goals.

### Effects of Goal Use on Induction: Blank Property Induction Task

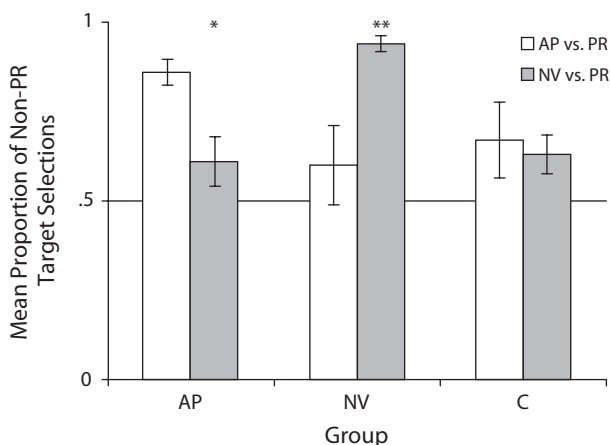
To test whether the pursuit of goals during knowledge acquisition affects the way that domain knowledge is used, the

**Table 1**  
Correlations Between the Ideal Matrices for the Pictorially Represented, Avoid Predator, and Nutritional Value Features and the Initial and Final Distance Matrices for Each Group in Experiment 1

Structure	Group		
	AP	NV	C
Initial Sorts (Day 1)			
PRi	.96	.99	.93
APi	.40	.13	.28
NVi	-.08	-.17	-.01
Final Sorts (Day 4)			
PRi	.35	-.11	.51
APi	.95	.22	.67
NVi	.15	.99	.53

Note—PRi, pictorially represented ideal; APi, avoid predators ideal; NVi, nutritional value ideal.

participants were presented with a blank property induction task. The participants' choices in each of the two trial types were scored according to whether they were consistent with the use of nonpictorial commonalities between the base and the target. Thus, for AP versus PR trials, the participant's choice was scored as a 1 if the AP target was chosen and as a 0 if the PR target was chosen. In NV versus PR trials, NV choices were scored as a 1 and PR choices as a 0. Figure 2 displays the mean proportions of non-PR target selections in each trial type for Groups AP, NV, and C. It appears that Groups AP and NV were most likely to generalize the novel blank property to a goal-related target when they were given the opportunity to do so—that is, in AP versus PR trials for Group AP and in NV versus PR trials for Group NV. In contrast, Group C appears to have made roughly the same number of non-PR target selections in each trial type. This pattern of results is supported by a 3 (group)  $\times$  2 (trial type: AP vs. PR and NV vs. PR) ANOVA on the mean proportions of non-PR target selections. The analysis revealed a significant group  $\times$  trial type interaction [ $F(2,18) = 12.18$ ,  $MS_e = 0.03$ ,  $p < .01$ ]. Follow-up analyses on this interaction compared the mean proportions of non-PR target selections in each trial type for each group. These follow-up analyses revealed that Group AP made more non-PR target selections in AP versus PR trials [ $F(1,18) = 8.00$ ,  $MS_e = 0.03$ ,  $p < .05$ ], Group NV made more non-PR selections in NV versus PR trials [ $F(1,18) = 15.96$ ,  $MS_e = 0.03$ ,  $p < .01$ ], and Group C made about the same proportion of non-PR selections in each trial type [ $F(1,18) = 0.23$ ,  $MS_e = 0.03$ , n.s.]. These results suggest that similarity with respect to goal-relatedness served as a basis for Groups AP and NV's inductions. For trials on which goal-relevant commonalities distinguished the two targets, members of Groups AP and NV were likely to select the non-PR target. When goal-relevant commonalities could *not* be used to distinguish the two targets, these participants were less likely to select the non-PR target, making about the same proportion of non-PR selections as did Group C.



**Figure 2.** Mean proportions of non-pictorially-represented (PR) target selections in the blank property induction task for the avoid predators (AP), nutritional value (NV), and control (C) groups as a function of trial type ( $\pm$  SE). \* $p < .05$ . \*\* $p < .01$ .

At the group level, the participants in the AP and NV conditions were found to use goal-related features across sorting and induction. We examined the data *within* groups to determine whether the participants who sorted more closely to their group's ideal also made inductions that were more consistent with the use of goal-related features. For the participants in Group AP, the association between correlation with APi and mean proportion of non-PR choices in AP versus PR trials was slightly negative [ $r(7) = -.10$ , n.s.]. For the participants in Group NV, the association between correlation with NVi and mean proportion of non-PR choices in NV versus PR trials was also negative [ $r(7) = -.22$ , n.s.]. Thus, there is no evidence that, within groups, bias toward goal relatedness in sorting is associated with bias toward goal relatedness in induction (in fact, there is some evidence to the contrary). These analyses, however, suffer from two severe limitations: (1) They were based on samples of only 7 individuals, and (2) the range of each variable (correlation with ideal and proportion of non-PR selections) was *highly* restricted. Thus, even if there were consistency across tasks at the individual level, the lack of variability within each group would prevent us from detecting it.

## Discussion

In Experiment 1, we sought to examine how the use of goals in an unsupervised learning environment affects the organization of domain knowledge and the application of this knowledge. We hypothesized that goal use would lead to a reliance on goal-relevant features in both the categorization and the induction tasks. For categorization, idealized structures based on goal-relevant properties provided a good fit to Groups AP and NV's categorizations. This implies that the participants did not merely distinguish "good" from "bad" exemplars with respect to their goal; they were sensitive to the particular features that contributed to these evaluations and categorized on the basis of exact matches of these features. The participants in the AP and NV conditions also based their property inductions on goal-relevant feature commonalities. Group C, who used all of the feature types about equally in their goal-directed task, did not evidence a bias toward any particular feature type in categorization or induction.

The results across the categorization and induction tasks suggest that the participants structured their knowledge of the experimental stimuli around ideal features—those related to the goal they had in evaluating the stimuli. This finding corroborates similar suggestions from correlational studies about the influence of goal use—for instance, Medin et al.'s (1997) study of tree experts. Indeed, the effects of goal use in the present experiment are relatively clear-cut (no pun intended), since the correlations between different feature types were strictly controlled. Our findings are also consistent with previous research on the effects of interacting with stimuli prior to categorization (e.g., Ross, 1996, 2000). They add to this prior research by demonstrating that goal use may affect categorization in the absence of instruction about category structure. Finally, the results of this experiment align with Love's (2003) finding that participants' learning modes (i.e., *how* they interacted with

exemplars) affected their evaluations of category membership for simple one-dimensional stimuli. The results of the present study show a similar effect, using multidimensional stimuli and a hierarchical categorization measure.

Experiment 1 showed that participants are biased toward goal-related information across different tasks, categorization and induction. Will this bias extend to a context in which goal-unrelated information is task relevant? Some research suggests that extended goal use may lead to a general bias toward goal-relevant information. Hashem, Chi, and Friedman (2003), for example, found that medical experts were more likely to overdiagnose diseases from their own area of specialization (e.g., a cardiologist who favors a diagnosis of heart disease over other likely diagnoses). Similarly, in an experimental study, Ross (2000) found that the use of certain information to prescribe medical treatments led to a bias to use this information in later tasks, such as disease diagnosis.

We examined the limitations of participants' goal-related biases in Experiment 2 by presenting them with an induction task that included meaningful properties, rather than blank properties. Previous research has shown that people may use different features as a basis for induction, depending on the meaning of the projected property. Heit and Rubinstein (1994), for example, found that biological class inclusion was used for induction of novel *anatomical* properties (e.g., base and target are *mammals*) but that, when the novel property was *behavioral*, it was more likely to be projected on the basis of behavioral similarity (e.g., base and target are *predators*). Our question was whether participants who completed a phase of goal-directed interactions would use property-related features as a basis for induction or whether they would continue to rely on goal-related information. In other words, would they display a general bias toward goal-relevant information, or would they apply their knowledge flexibly?

## EXPERIMENT 2

In Experiment 2, we used essentially the same tasks and stimuli as those in Experiment 1. Our main extension was to examine property induction with nonblank (*meaningful*) properties—unfamiliar, novel properties related to either avoiding predators or nutritional value. This allowed us to examine whether the participants would base their inductions on goal-relevant commonalities, as was found in Experiment 1, or on commonalities related to the meaning of the property.

### Method

#### Participants

Twenty-seven undergraduates from the University of Illinois at Chicago participated to fulfill a course requirement. The participants were run individually, and 9 participants were randomly assigned to each of three conditions.

#### Materials and Tasks

The same materials and tasks as those in Experiment 1 were used in Experiment 2, with the exception that the blank property induction task was replaced with the nonblank property induction task described below.

**Nonblank property induction task.** On each trial of this task, the participant was informed that a given creature was found to possess a novel property: *the ability to detect predators from far away* (AP), *possessing high nutrient content* (NV), or *having abnormal sarcaloid levels* (blank). The nonblank AP property was intended to be straightforwardly related to the AP features from the stimulus set, and the nonblank NV property was intended to be straightforwardly related to the NV features. The blank property was intended to appear equally related to the AP and NV features.

The participant's task was to select which of two target creatures was most likely to also have the novel property that was attributed to the base creature. The base creature had novel pictorial features—for example, swirled body markings and novel facial features—so there was minimal pictorial feature overlap between the base and the targets. This base creature possessed familiar written features, three AP and three NV, which were distributed in line with the frequencies listed in Appendix A. In all the trials, one of the targets had more AP features in common with the base (AP target), and the other had more NV features in common with the base (NV target). Thus, the participant could project the novel property on the basis of either AP or NV feature commonalities between the base and the target. On all the trials, AP competed with NV. Each trial was presented on a computer screen. The original item was presented for 10 sec and was then replaced by two targets that were displayed until the participant made a selection. Following the response, a blank screen was displayed for 2 sec, followed by the next trial. The participant was presented with an equal number of trials in which AP, NV, and blank properties were attributed to the base creature (12 of each). Trial order was randomized to create a single task list that was administered to each participant.

#### Procedure

The same procedure as that in Experiment 1 was followed, with the exception that on Day 4, the participants in this experiment completed the nonblank property induction task.

## Results

### Effects of Goal Use on Categorization: Hierarchical Sorting Task

As in Experiment 1, distance matrices were computed for each group's initial and final sorts and were correlated with the matrices derived from the ideal structures—PR<sub>i</sub>, AP<sub>i</sub>, and NAP<sub>i</sub>. Table 2 displays these correlations. All three groups initially favored the use of PR features for sorting on Day 1 but displayed different tendencies in their final sorts on Day 4, with Groups AP and NV showing close alignment with their respective ideals (AP<sub>i</sub> and NV<sub>i</sub>) and Group C showing a modest correlation with each ideal. Analyses were once again conducted to determine which ideal structure, AP<sub>i</sub> or NV<sub>i</sub>, provided a better fit to each group's final sort. These analyses revealed that AP<sub>i</sub> provided a significantly better fit than did NV<sub>i</sub> for Group AP's final sort [ $r(120) = .45, p < .01$ ], NV<sub>i</sub> provided a better fit for Group NV's final sort [ $r(120) = .48, p < .01$ ], and AP<sub>i</sub> and NV<sub>i</sub> fit Group C's final sort equally well [ $r(120) = .15, n.s.$ ]. Altogether, the results of the hierarchical sorting task essentially replicated the results from Experiment 1. Again, Groups AP and NV formed categories in close alignment with their respective ideal category structures.

Up to this point, each group's categorization performance has been considered only in the aggregate. It is particularly of interest to examine Group C's categorizations at the individual level. Is the observed pattern of

**Table 2**  
**Correlations Between the Ideal Matrices for the Pictorially Represented, Avoid Predator, and Nutritional Value Features and the Initial and Final Distance Matrices for Each Group in Experiment 2**

Structure	Group		
	AP	NV	C
Initial Sorts (Day 1)			
PRi	.99	.97	.99
APi	.18	.22	.20
NVi	-.14	-.14	-.11
Final Sorts (Day 4)			
PRi	.18	-.04	.40
APi	.94	.34	.69
NVi	.35	.97	.50

Note—PRi, pictorially represented ideal; APi, avoid predators ideal; NVi, nutritional value ideal.

correlations (i.e., moderate correlations with each ideal) representative of the performance of the individual participants in Group C, or could it be a product of averaging across participants who had a relatively pure preference for different feature types? To address this question, we plotted the correlations between each participant's final distance matrix and the AP and NV ideal matrices. Figure 3 displays these data for all the participants across Experiments 1 and 2 ( $n = 16$ ). As was expected, the participants in Groups AP and NV formed two distinct clusters, one highly correlated with APi and the other highly correlated with NVi. Note that 4 participants in Group AP and 7 in Group NV evidenced perfect correlations with their respective ideals; thus, their data points overlap.

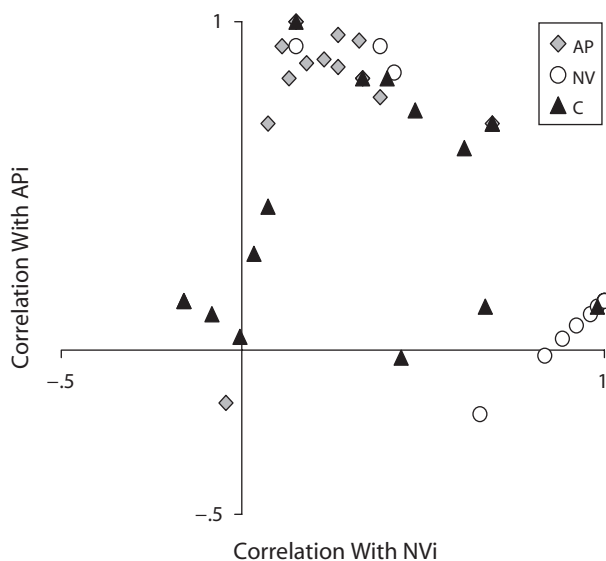
If the participants in Group C had a relatively pure preference for either AP or NV features in their sorting, they ought to fall within one of these clusters. However, it is clear from Figure 3 that this was not the case. The participants in Group C are distributed far less uniformly than are those in the other two groups, with some evidencing relatively low correlations with each ideal, whereas others are relatively high. Only a couple of participants in Group C show a clear bias toward one of the two ideals. Thus, we can conclude that the mean pattern of results for each group does not distort the performance at the individual level. This lends further support to our conclusion that forming categories around a particular feature type depended on whether this type was used exclusively in goal-directed activity prior to categorization.

**Effects of Goal Use on Induction:  
 Nonblank Property Induction Task**

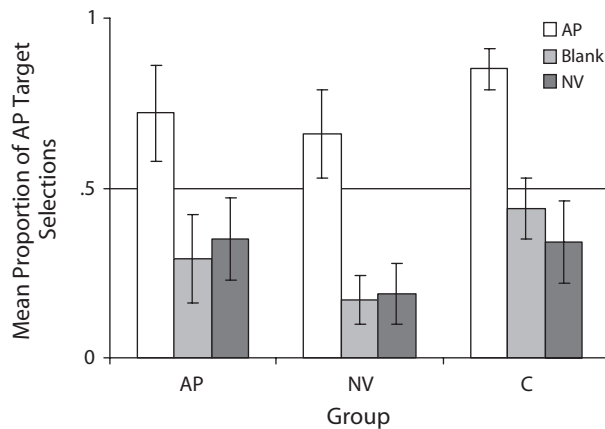
This task had three different trial types, which differed in terms of whether an AP, NV, or blank property was attributed to the base creature. On each trial, an AP target competed against an NV target. We examined the mean proportion of AP target selections that the participants made in each trial type, although the same results would be obtained using mean proportion of NV target selections as the dependent variable. Figure 4 displays the mean proportion of AP target selections by each group for each of the three trial types—AP, NV, and blank.

If each group made their target selections on the basis of goal-relevant feature commonalities, as was clearly the case for Groups AP and NV in Experiment 1, Group AP ought to have made a *high* proportion of AP target selections regardless of trial type, Group NV ought to have made a consistently *low* proportion of AP selections, and Group C could be expected to fall somewhere in between. Clearly, this was not the case. In fact, the meaning of the property seems to have played a significant role in each group's inductions. All the groups appear to have favored AP targets in AP feature trials but selected NV targets in NV and blank feature trials (thus, the lower proportion of AP target selections). This observed pattern of results is supported by a 3 (group)  $\times$  3 (trial type: AP, NV, or blank) ANOVA on the mean proportions of AP target selections. This analysis revealed main effects of group [ $F(2,24) = 3.93, MS_e = 0.07, p < .05$ ] and trial type [ $F(2,48) = 13.64, MS_e = 0.13, p < .01$ ], but the group  $\times$  feature type interaction was not significant [ $F(4,48) = 0.15, MS_e = 0.13, n.s.$ ]. Follow-up analyses on the main effect of trial type revealed that the participants made a significantly higher proportion of AP target selections in AP feature trials than in NV and blank trials [ $F(1,48) = 8.87, MS_e = 0.13, p < .01$ ]. Follow-up analyses on the main effect of group revealed that Group NV made a lower proportion of AP target selections than did Group C [ $t(16) = 2.90, p < .05$ ].

Overall, when the participants were asked to project an AP-related property to either an AP or an NV target, they were likely to select the AP target. When asked to project an NV-related or blank property, they were *less* likely to select the AP target—that is, *more* likely to select the NV target. This pattern of results was observed across groups. However, the groups were found to differ in the proportion of AP target selections they made overall. Group NV was



**Figure 3.** Correlations between each participant's final distance matrix and the avoid predators (AP) and nutritional value (NV) ideal matrices. The data from both Experiments 1 and 2 are displayed.



**Figure 4.** Mean proportions of avoid predators (AP) target selections in the nonblank property induction task for the AP, nutritional value (NV), and control (C) groups as a function of trial type ( $\pm SE$ ).

the least likely to select the AP target across trials; that is, they were the *most likely* to select the NV targets.

One unexpected result from the nonblank induction task is that the participants in each group projected the blank property, *abnormal sarcaloid levels*, much as they did the NV-related property. Perhaps the participants interpreted this property as being related to the NV features because it seems to refer to some internal, possibly biochemical substance. Thus, the property may not have been as “blank” as we had intended. In support of this conjecture, we presented each list of AP and NV feature meanings to 28 naive participants and asked them to indicate which list seemed to fit the AP, NV, and blank properties used in the nonblank induction task best. All 28 participants indicated that the NV property fit best with the NV feature meanings, and 26 out of 28 indicated that the AP property fit best with the AP list. As was suspected, 23 out of the 28 participants indicated that the blank property fit best with the NV feature meanings [ $\chi^2(1) = 11.57$ ,  $p < .01$ ]. Thus, it appears that the blank property was, in fact, biased toward one property type. Assuming that the participants in Experiment 2 were also inclined to relate the blank property to NV features, their bias toward NV feature commonalities on the blank property trials seems quite reasonable.

### Discussion

Experiment 2 replicated the categorization results from the first experiment, finding that the participants in Groups AP and NV sorted the stimuli on the basis of goal-relevant commonalities. Whereas Experiment 1 showed that the use of certain features in goal-directed activities led to their use for the projection of blank properties, Experiment 2 showed that the projection of nonblank (unfamiliar but meaningful) properties was not limited to goal-relevant commonalities. Thus, the participants were capable of applying their knowledge flexibly, using the information that was most salient within the context of the task, regardless of whether it was goal related or unrelated.

Together, Experiments 1 and 2 suggest that goal-directed interactions with stimuli produce knowledge that is organized around goal-related features. What is it about goal-directed learning that leads to these effects? A reasonable possibility is that the participants in the experimental conditions organized their knowledge around goals because this structure seemed most meaningful to them. For the participants in the control groups, there was little consensus on the properties related to their goal; thus, a single ideal organizational structure did not emerge. There is, however, an alternative explanation for our results. In the experimental conditions, goal-directed learning required the participants to selectively attend to particular subsets of goal-related information (the memory component of the interaction task) and to evaluate category members with respect to their goal (the evaluation component). Perhaps the effects of goal-directed learning were simply a by-product of the participants’ attending to and using certain information and did not depend on the presence of a meaningful goal. If this was the case, simply directing the participants to attend to and use certain subsets of features would have led them to organize their knowledge around these features. Although we suspect that in everyday life, selectively attending to and using certain information rarely occurs in the absence of a meaningful goal, it is possible to tease apart *mere use* from the *meaningful use* that the participants performed in the first two experiments. This was the purpose of Experiment 3.

### EXPERIMENT 3

In this experiment, we examined the effects of the *mere use* of certain information. In each of two conditions, the participants were required to use a certain subset of stimulus features, but they were not given a meaningful goal as the basis for doing so. One group of participants was asked to monitor the creatures’ *R-type* features, whereas the other was asked to monitor the *B-type* features. These subsets of features corresponded to the AP-related and NV-related features, respectively. Goal-relevant information was explicitly revealed to the participants by coloring each creature’s written features either red or blue. The red features corresponded to the R-type/AP features, and the blue corresponded to the B-type/NV features. The participants therefore used exactly the same information as that used by the participants in Groups AP and NV in the first two experiments. The same tasks as those in Experiment 1 were presented, with slight modifications to accommodate the novel goals.

Our question was whether the participants would subsequently organize their knowledge around the features that they used, similar to our findings for the meaningful use conditions in the previous experiments. If they did, this would suggest that the emergence of goal-related knowledge structures might be a by-product of simply attending to and using certain stimulus features and did not depend on the presence of a meaningful goal. If the participants in the mere use conditions did not appear to organize their knowledge around the features that they used, this would suggest that the meaningfulness of the goal was critical to such effects.

## Method

### Participants

Fourteen undergraduates from the University of Illinois at Chicago participated to fulfill a course requirement. The participants were run individually, and 7 participants were randomly assigned to each of two conditions.

### Materials and Tasks

The same materials and tasks as those in Experiment 1 were used in Experiment 3, except that different goals were presented to the participants and the interaction task was modified to reflect these different goals. Note that the participants performed the blank induction task, as in the first experiment.

**Goals and background story.** One group of participants, Group Red, was told that they had been hired as consultants for a large corporation. They were told that the corporation was interested in documenting the creatures' R-type features and that these features would be indicated in red. The other group of participants, Group Blue, was told a similar background story, except that they were required to document the creatures' B-type (blue) features.

**Modified interaction task.** During the modified task, the AP- and NV-related written features were colored red and blue, respectively. The written features were colored only on the stimuli, not in the feature checklists that the participants completed for the memory component of the task. The written features were not colored for any of the other tasks. As in the previous experiments, the interaction task included a memory component, in which the participants were asked to indicate the R-type or B-type features of a recently presented creature. There was also an evaluation component, in which the participants were asked (1) whether they thought it would be difficult to remember the creature's R-type (or B-type) features later on, (2) why they thought it would be difficult or not, and (3) how difficult it would be, on a 1–10 scale. The evaluation component, like the memory component, was intended to keep the participants focused on goal-related information in completing each task component. The participants saw the same items in the same order as in the previous interaction tasks.

### Procedure

The same procedure as that in Experiment 1 was followed, except that the participants were given different goals and background stories and they performed the modified interaction task.

## Results

### Effects of Mere Use on Categorization: Hierarchical Sorting Task

As in Experiments 1 and 2, distance matrices were computed for each group's initial and final sorts and were correlated with the matrices derived from the ideal structures—PRi, APi, and NVi. Table 3 displays these correlations. As in the previous experiments, both groups initially favored the use of PR features for sorting on Day 1. Unlike in our prior findings for Groups AP and NV, however, the participants in the mere use conditions did not evidence a marked shift toward goal-relevant organizations in their final sorts. Group Red (AP) evidenced strong correlations with both the APi and PRi structures. Indeed, these two structures fit Group Red's final sort equally well [ $r(120) = .15$ , n.s.]. Group Blue (NV) evidenced moderate correlations with all three ideal structures, much like the control groups in the previous experiments. This group's final sort corresponded most closely with the APi and NVi structures; however, these two structures fit Group Blue's final sort equally well [ $r(120) = .07$ , n.s.].

From inspecting the data in Table 3, there is some indication that mere use affected the information that the par-

**Table 3**  
Correlations Between the Ideal Matrices for the Pictorially Represented, Avoid Predator, and Nutritional Value Features and the Initial and Final Distance Matrices for Each Group in Experiment 3

Structure	Group	
	Red (AP)	Blue (NV)
	Initial Sorts (Day 1)	
PRi	.98	.99
APi	.18	.14
NVi	-.14	-.19
	Final Sorts (Day 4)	
PRi	.63	.35
APi	.82	.62
NVi	.07	.71

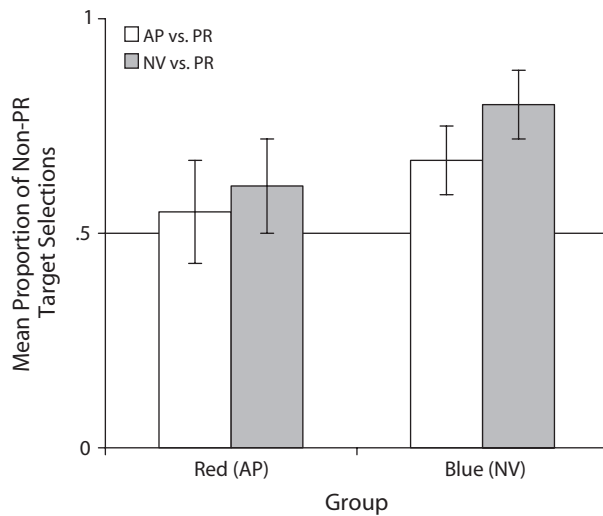
Note—PRi, pictorially represented ideal; APi, avoid predators ideal; NVi, nutritional value ideal.

ticipants used in their final sorts. The final sort for Group Red (AP) was most highly correlated with APi, whereas the final sort for Group Blue (NV) correlated highest with NVi. Inspection of the individual data revealed that these trends were driven mainly by two members within each group, whose sorts were nearly identical to their group's ideal. Perhaps these participants noticed the commonalities underlying the different colored features or simply responded more strongly to the mere use manipulation. At any rate, goal-related biases were not observed for the majority of the participants. Overall, the mere use of particular features, unlike meaningful use, did not result in a significant shift toward a goal-related organization.

### Effects of Mere Use on Induction: Blank Property Induction Task

In this task, as in Experiment 1, the participants were asked to project a novel blank property from a base item to one of two possible target items. In two trial types, either AP and PR targets or NV and PR targets competed. For each group, we computed the mean proportion of non-PR selections for each trial type. If the mere use of certain features leads participants to use these features for induction, we would expect Group Red (AP) to make a higher proportion of non-PR selections in AP versus PR trials and Group Blue (NV) to make a higher proportion of non-PR selections in NV versus PR trials. Figure 5 displays the mean proportions of non-PR target selections in each trial type for each group.

Unlike in the meaningful use groups, the mere use groups do not appear to differ in their selection of non-PR targets across trial types. It does appear that the participants made slightly more non-PR selections in the NV versus PR trials. It also appears that the participants in Group Blue (NV) were more likely to select the non-PR target than were the participants in Group Red (AP). To explore these results further, we conducted a 2 (group)  $\times$  2 (trial type: AP vs. PR and NV vs. PR) ANOVA on the mean proportions of non-PR target selections. This analysis, however, revealed no main effects of group [ $F(1,12) = 2.68$ ,  $MS_e = 0.11$ , n.s.] or trial type [ $F(1,12) = 1.58$ ,  $MS_e = 0.02$ , n.s.]. There was no interaction between group and trial type



**Figure 5.** Mean proportions of non-pictorially-represented (PR) target selections in the blank property induction task for Groups Red (avoid predators [AP]) and Blue (nutritional value [NV]) as a function of trial type ( $\pm SE$ ).

[ $F(1,12) = 0.35$ ,  $MS_e = 0.02$ , n.s.]. Altogether, we find no evidence that mere use of particular stimulus features led the participants to use these features as a basis for the induction of a novel property. The participants made about the same proportion of non-PR selections, regardless of their goal condition and regardless of the trial type.

When categorization was considered at the individual level, 2 participants in each condition were found to have sorted the stimuli closely to their respective ideal. Did these same participants favor goal-related properties as a basis for inference? For the participants in Group Red (AP), one of these ideal sorters made slightly more non-PR selections in AP versus PR trials, whereas the other made more non-PR selections in NV versus PR trials. For the ideal sorters in Group Blue (NV), both made more non-PR selections in NV versus PR trials, but the difference was pronounced for only 1 of these participants and was slight for the other. Thus, although all 4 individuals showed a strong bias toward ideals in their sorts, only 1 participant evidenced a comparable bias in their inductions. Perhaps this ideal sorter noticed the meaningful relations between the features that they had used, whereas the other 3 did not.

### Discussion

The main purpose of Experiment 3 was to assess the importance of having a meaningful goal in the context of learning about a novel domain. In particular, we were interested in whether participants who were simply directed to attend to and use certain features (mere use) would then rely on these features for organizing their domain knowledge. Our results revealed that the mere use of certain information did not result in the effects that we observed in the meaningful use conditions in the previous experiments. On average, the participants in the mere use conditions did not evidence a clear shift toward a goal-related organiza-

tion in their final sorting of the stimuli, nor did they favor the use of goal-related information in their induction of a novel property. These results suggest that the presence of a meaningful goal—that is, one that links the goal and the selected features in a nonarbitrary fashion—plays an important role in the emergence of goal-related knowledge structures. We will elaborate further on this point in the General Discussion section that follows.

### GENERAL DISCUSSION

Altogether, our experiments showed that participants who pursued a meaningful goal were likely to organize items into goal-based categories. In general, these effects were not observed when the participants were given a goal that did not require the use of particular stimulus features (the control conditions in the first two experiments) or when the goal was nonmeaningful (the mere use conditions in the third experiment). The participants in the meaningful use conditions also used goal-relevant information as a basis for the induction of novel properties; however, this was found to depend on the property's meaning. When induction involved nonblank (meaningful) properties, the participants tended to base their inductions on property-related features, regardless of whether they were goal related or unrelated. Only when the property was not straightforwardly related to certain category features (i.e., a blank property) did the participants appear to base their inductive generalizations on goal-related commonalities.

Our findings from the hierarchical sorting task align well with those from previous research involving expert populations (e.g., Burnett et al., 2005; Medin et al., 1997; Shafto & Coley, 2003). Like expert specialists who devote a great deal of time engaging in goal-related activities in their domain, the participants in the meaningful use conditions were found to organize their domain knowledge around goals. Unlike previous correlational research, however, our experiments establish a causal relation between goal use and knowledge organization. Our findings also imply that goals play a role in knowledge organization beyond such factors as the frequency of exposure to stimuli and the use of certain stimulus features. Indeed, the participants in the mere use conditions were exposed to the same stimuli and used the same information as did the participants in the meaningful use conditions, but they did not appear to develop goal-related organizations of the stimuli. We will discuss how goals may affect knowledge acquisition and organization in further detail below.

Our induction findings also align well with those from previous research demonstrating that individuals with a wealth of knowledge in a given domain, such as experts with highly specialized knowledge or laypeople with common knowledge of a familiar domain, draw on different aspects of their knowledge in different contexts of inference. For instance, although different tree experts may categorize trees into goal-related categories (Medin et al., 1997), they may rely on their knowledge of causal and ecological relations to project properties (Proffitt, Medin, & Coley, 2000; Shafto & Coley, 2003). As Medin, Coley,

Storms, and Hayes (2003) suggested, category-based induction may draw on the most salient or relevant information at hand. In our experiments, goal-related properties may have been most salient when the novel property could not be associated with other known properties—that is, when it was blank or meaningless. When the property could be associated with other known features, these other features were then used.

In what contexts might we observe limitations to the flexible use of knowledge; in other words, when might the learner display goal-related biases? One important factor is the nature of the task that is presented to participants. Previous research on the effects of goal use has shown evidence of goal-related biases in such tasks as feature generation and ratings of feature frequency (Ross, 1997, 2000). Other studies have shown that experts are biased toward ideals in their ratings of exemplar *goodness* or typicality (Burnett et al., 2005; Lynch et al., 2000). It would be of interest to incorporate such tasks into the present paradigm. Comparisons between the mere and the meaningful use conditions might be especially informative, since they could address whether it is lower level selection of features or the presence of meaningful knowledge structures that leads to goal-related biases. Knowledge flexibility may also be affected by the way in which the participants learn about the domain—that is, their mode of learning. As Markman and Ross (2003) argued, the way in which category information is learned has important consequences for how it is represented and used. It would be of interest to extend the present work by comparing different modes of learning, such as supervised versus unsupervised learning or classification versus inference learning. Perhaps, for example, supervised learning would lead participants to appear more biased toward goal-relevant information, because category boundaries are perceived as more rigid. By examining performance across different tasks and different modes of learning, we could further explore the limitations of flexibility and the causes of these limitations.

### The Function of a Goal in Knowledge Acquisition and Organization

In the context of learning about a novel domain, the function of a meaningful goal may be twofold. First, in light of the learner's background knowledge, the goal provides a basis for intelligently selecting and ignoring information. As Barsalou (1995) suggested, the storage of goal-relevant information may be a side effect of this selective process, not necessarily an intended outcome. Second, the goal can provide coherence to the members of an otherwise nonsensical category, such as *things to take out of a house in the event of a fire* (Barsalou, 1983). Our use of the term *coherence* aligns closely with Murphy and Medin's (1985) usage, in which a coherent category is a grouping of objects that *makes sense* to the perceiver. For this coherence to be appreciated, an individual must possess knowledge of the goal-related features of category members; it is this background knowledge that ultimately determines the sensibleness of the grouping. If, for example, a nonexpert were shown the goal-related categories

formed by one of the tree experts in Medin et al.'s (1997) study, the organizational structure would not seem sensible; it might, in fact, appear quite mysterious. To the appropriate expert, however, the organization would appear coherent, because he or she would be aware of the relevant properties that category members share. Here, goal-relevant knowledge plays a role similar to that of a theme in a text passage. Unless the theme is known, a text may seem to be no more than a curious collection of unrelated sentences (e.g., Bransford & Johnson, 1973).

In our first two experiments, the presence of a meaningful goal enabled the participants to select goal-relevant information and provided a basis for organizing the stimuli into goal-related categories. Thus, both functions of a goal, selection and coherence, were likely at play. When the connection between the participants' goals and the selected features was arbitrary, as in the mere use conditions, the effects on knowledge organization and induction largely disappeared. This suggests that the coherence provided by a goal is critical to the formation of goal-related categories. Without the "conceptual glue" provided by a meaningful goal, a goal-based organization may not emerge.

Our results, however, do not completely rule out the importance of feature selection in the formation of goal-related categories. It could be argued that the meaningful use conditions required more effortful feature selections as the participants learned about the domain. Indeed, during the interaction task, the participants in the meaningful use conditions selected features on the basis of their meanings, whereas the participants in the mere use conditions relied on color. To more thoroughly determine the contributions of coherence and selection, it would be desirable to control for the demands of selection during the goal-directed learning phase. This, however, is a difficult problem: How can processing effort be equated between a condition in which selection is based on feature meaning and one in which selection is based on some other aspect of the features? If selection were meaning-based in both conditions, we would potentially be introducing a new, meaningful goal to the learning context. It may be more tractable to devise a condition in which the goal is meaningful but goal-relevant features are indicated by a superficial characteristic, such as color. Thus, the meaningfulness of the goal would be controlled, while the demands of selection would be varied. This might allow us further insight into the contribution of the feature selection process itself. Currently, what we can conclude is that the meaningful connection between a goal and its related features is of great importance to the development of goal-based knowledge. This appears to be due mainly to the coherence that this goal provides to a category, but it may also be due, in part, to the "deeper," more effortful processing required to select features under a meaningful goal.

A final question to consider is the role of interacting with stimuli. After a goal is defined, what is the additional impact of *using* this goal? Although our research did not separate the mere presence of a meaningful goal from its use, there is reason to believe that actually using a goal does have an additional influence on category knowledge.

Ross and Warren (2002), for example, found that goal use leads to an appreciation of abstract category relations that would otherwise go unnoticed. Other research by Ross (e.g., 1997, 2000) has shown that goal use affects people's judgments about the frequency and distribution of stimulus features. Although these studies suggest ways in which goal use may be unique, it would be of interest to contrast goal presence with goal use in a single experiment. Perhaps the methodology presented in this article could be extended to explore this issue.

### Relating This Research to Real-World Learning

Outside of the laboratory, goal-directed learning generally operates in the presence of a meaningful goal, one that links the goal and the selected features in a nonarbitrary fashion. Thus, there is reason to think that our results from the meaningful use conditions will be generalizable to real-world learning contexts. Yet, in extrapolating our results, it is important to recognize that goal-directed learning may be far more complex outside of the laboratory. With our stimuli, all the features of each item (pictorial and written) were available to the participant at once. In the real world, goal-relevant features must often be retrieved from representations stored in memory, rather than merely being recognized in the presented stimulus. The features that become available can be biased by the context of evaluation (Barsalou, 1982; Tversky & Kahneman, 1974), and even interpreting the available features may be context dependent (Medin, Goldstone, & Gentner, 1993; Wisniewski & Medin, 1994). Another consideration is that, whereas the written features in our experiments were either positive or negative with respect to a particular goal, features of real-world stimuli may vary in their idealness on a continuous scale. For instance, rather than being either high or low in fat, foods may vary in fat content on a continuous scale, from 0 g on upward. Previous research has shown that people are sensitive to the continuous nature of ideals in their evaluations of typicality for goal-based categories (Barsalou, 1985), so they may show similar sensitivity in knowledge organization. Finally, the stimuli used in these experiments were designed so that they had either three positive or three negative features with respect to each goal. Real-world stimuli, in contrast, may have both positive *and* negative features. Thus, evaluating their idealness may require the use of a more complicated cost-benefit analysis. All of these complexities highlight the need for further research into the effects of goal use on knowledge acquisition.

Besides the learning itself, it is important to consider the ecological validity of our measures of knowledge organization and use. Do our sorting results imply that the participants developed the particular category structures that they produced during the task? It may be the case that they did, but it is also possible that these structures were assembled ad hoc. Although our results do imply that the participants prioritized goal-relevant information, they do not imply the presence of goal-based categories in long-term memory. It would be of interest to address this limitation in future research. A similar consideration can be raised about our findings in the induction task. Here, too,

the participants could inspect all the features of each target item. Thus, they did not have to rely on their memory for these features, as is required when only a category label is provided. It is interesting to ask whether the participants in our study would have used property-related commonalities for induction if they had been required to retrieve information about the base and target items from memory. Perhaps both blank and nonblank properties would be projected on the basis of goal-related commonalities, since goal-relevant information may be better encoded. This is also a question for future research.

In closing, we would like to emphasize that continued research is required to explore the effects of specialization in expert populations and to integrate this research with relevant experimental findings. Some recent studies have documented interesting differences in knowledge and reasoning between expert specialists in domains such as aquatic systems (Hmelo-Silver & Pfeffer, 2004), medical diagnosis (Hashem et al., 2003), and tree life (Proffitt et al., 2000). Other research has explored cross-cultural differences in biological categorization and reasoning (e.g., Bailenson, Shum, Atran, Medin, & Coley, 2002; Medin, Ross, Atran, Burnett, & Blok, 2002). There also exists a fairly extensive body of research on artifact categorization, which bears some interesting relations to the research on goal-related categorization. For one, membership in an artifact category, such as *ashtray* or *spoon*, depends critically on the *intended* function of an object (e.g., Matan & Carey, 2001). Such attributions of intention may involve associations between the goals of an agent (the creator) and the manifest features of an artifact (the creation), analogous to those present in goal-related categories. Although it is beyond the scope of this article to explore this issue in full, comparing the function of goals in the acquisition and use of artifact categories would provide an interesting extension to this work. Altogether, these diverse lines of research may provide fertile ground for a comprehensive and interdisciplinary approach to understanding the relationship between goal-directed interactions and the development of knowledge across the continuum of expertise.

### AUTHOR NOTE

This research was supported by an NSERC PGS-A fellowship to B.D.J. Portions of this research are based on a thesis submitted by B.D.J. to the University of Illinois at Chicago, in partial fulfillment of the requirements for an M.A. degree. We thank Len Newman and Jim Pellegrino for helpful discussions about this research and Robert Ariel, Jeremy Eagles, Teresa Susmaras, and Himanshu Thakur for their assistance with running participants and entering data. Correspondence concerning this article should be addressed to B. D. Jee, Department of Psychology, 1007 Harrison St. (M/C 285), University of Illinois, Chicago, IL 60607 (e-mail: bendj@uic.edu).

### REFERENCES

- BAIENSON, J. N., SHUM, M. S., ATRAN, S., MEDIN, D. L., & COLEY, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition*, *84*, 1-53.
- BARSALOU, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82-93.
- BARSALOU, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211-227.
- BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instan-

- tion as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 629-654.
- BARSALOU, L. W. (1995). Storage side effects: Studying processing to understand learning. In A. Ram & D. Leake (Eds.), *Goal-driven learning* (pp. 407-419). Cambridge, MA: MIT Press, Bradford Books.
- BÉDARD, J., & CHI, M. T. H. (1992). Expertise. *Current Directions in Psychological Science*, **1**, 135-139.
- BRANSFORD, J. D., & JOHNSON, M. K. (1973). Consideration of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing* (pp. 383-438). New York: Academic Press.
- BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. (1956). *A study of thinking*. New York: Wiley.
- BURNETT, R. C., MEDIN, D. L., ROSS, N. O., & BLOK, S. V. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology*, **59**, 3-10.
- CHI, M. T. H., FELTOVICH, P., & GLASER, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121-152.
- HASHEM, A., CHI, M. T. H., & FRIEDMAN, C. P. (2003). Medical errors as a result of specialization. *Journal of Biomedical Informatics*, **36**, 61-69.
- HEIT, E., & RUBINSTEIN, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 411-422.
- HMELO-SILVER, C. E., & PFEFFER, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, **28**, 127-138.
- JEE, B. D., & WILEY, J. (2004). The influence of goal-directed activity on categorization and reasoning. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 630-635). Mahwah, NJ: Erlbaum.
- LOVE, B. C. (2003). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin & Review*, **10**, 190-197.
- LYNCH, E. B., COLEY, J. D., & MEDIN, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, **28**, 41-50.
- MARKMAN, A. B., & ROSS, B. H. (2003). Category use and category learning. *Psychonomic Bulletin*, **129**, 592-615.
- MATAN, A., & CAREY, S. (2001). Developmental changes within the core of artifact concepts. *Cognition*, **78**, 1-26.
- MEDIN, D. L., COLEY, J. D., STORMS, G., & HAYES, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, **10**, 517-532.
- MEDIN, D. L., GOLDSTONE, R. L., & GENTNER, D. (1993). Respects for similarity. *Psychological Review*, **100**, 254-278.
- MEDIN, D. L., LYNCH, E. B., COLEY, J. D., & ATRAN, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, **32**, 49-96.
- MEDIN, D. L., ROSS, N., ATRAN, S., BURNETT, R. C., & BLOK, S. V. (2002). Categorization and reasoning in relation to culture and expertise. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 1-41). New York: Academic Press.
- MEDIN, D. L., & SCHWANENFLUGEL, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1249-1265.
- MURPHY, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- MURPHY, G. L. (2003). Ecological validity and the study of concepts. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 1-41). New York: Academic Press.
- MURPHY, G. L., & MEDIN, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289-316.
- MURPHY, G. L., & WRIGHT, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 144-155.
- NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- OSHERSON, D. N., SMITH, E. E., WILKIE, O., LÓPEZ, A., & SHAFIR, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.
- PELLEGRINO, J. W., & HUBERT, L. J. (1982). The analysis of organization and structure in free recall. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 129-172). New York: Academic Press.
- PROFFITT, J. B., MEDIN, D. L., & COLEY, J. D. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 811-828.
- RIPS, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.
- ROSS, B. H. (1996). Category learning as problem solving. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 35, pp. 165-192). New York: Academic Press.
- ROSS, B. H. (1997). The use of categories affects classification. *Journal of Memory & Language*, **37**, 240-267.
- ROSS, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 743-757.
- ROSS, B. H. (2000). The effects of category use on learned categories. *Memory & Cognition*, **28**, 51-63.
- ROSS, B. H., & MURPHY, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, **38**, 495-553.
- ROSS, B. H., & WARREN, J. L. (2002). Learning abstract relations from using categories. *Memory & Cognition*, **30**, 657-665.
- SHAFTO, P., & COLEY, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naive similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 641-649.
- SHEPARD, R. N., HOVLAND, C. I., & JENKINS, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**(13, Whole No. 517).
- SMITH, M. U. (1990). Knowledge structures and the nature of expertise in classical genetics. *Cognition & Instruction*, **7**, 287-302.
- SMITH, M. U. (1992). Expertise and the organization of knowledge: Unexpected differences among genetic counselors, faculty, and students on problem categorization tasks. *Journal of Research in Science Teaching*, **29**, 179-205.
- TANAKA, J. W., & TAYLOR, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, **23**, 457-482.
- TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124-1131.
- WISNIEWSKI, E. J., & MEDIN, D. L. (1994). On the interaction between theory and data in concept learning. *Cognitive Science*, **18**, 221-281.
- YAMAUCHI, T., LOVE, B. C., & MARKMAN, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 776-795.
- YAMAUCHI, T., & MARKMAN, A. B. (1998). Category learning by inference and classification. *Journal of Memory & Language*, **39**, 124-148.

## NOTE

1. The stimuli used in the present experiment differed from those used in Jee and Wiley (2004) in terms of the feature labels. The feature labels used in the previous work were designed to be unrelated to their meanings. This made the meanings more difficult to learn overall. In the present experiment, we designed the labels to be straightforwardly related to their meanings. This allowed us to streamline the procedure, since less time would be devoted to learning the feature meanings.

**APPENDIX A**  
**Pictorially Represented Features in the Stimuli**

Written Description of Feature	Frequency	Stimuli With Feature
Eyebrows	2/16	1, 4
Horns	2/16	5, 8
Tail	2/16	9, 12
Cleft chin	2/16	13, 16
Large ear	2/16	2, 3
Spines on back	2/16	6, 7
Round nose	2/16	10, 11
Cheekbones	2/16	14, 15
Cone-shaped head	4/16	1, 4, 5, 8
Feather on head	4/16	9, 12, 13, 16
Antennae on head	4/16	2, 3, 6, 7
Hair on head	4/16	10, 11, 14, 15
Spots on body	8/16	1, 4, 5, 8, 9, 12, 13, 16
Stripes on body	8/16	2, 3, 6, 7, 10, 11, 14, 15

**APPENDIX B**  
**Feature Labels and Meanings in the Stimuli**

Feature Label	Meaning	Frequency	Stimuli With Feature
Features Related to Avoiding Predators			
Camo arch	camouflage capability	2/16	1, 2
Wilnest in gut	nest-building ability	2/16	3, 4
Izzdedd cells	able to play dead	2/16	5, 6
Hidigg lining	can dig holes for hiding	2/16	7, 8
Nocamo arch	<i>no</i> camouflage capability	2/16	9, 10
Nowilnest in gut	<i>no</i> nest-building ability	2/16	11, 12
Noizzdedd cells	<i>not</i> able to play dead	2/16	13, 14
Nohidigg lining	<i>cannot</i> dig holes for hiding	2/16	15, 16
Kwickru in blood	able to run quickly	4/16	1, 2, 3, 4
Sigmake vessel	can signal for help	4/16	5, 6, 7, 8
Nokwickru in blood	<i>not</i> able to run quickly	4/16	9, 10, 11, 12
Nosigmake vessel	<i>cannot</i> signal for help	4/16	13, 14, 15, 16
Divin membrane	can dive underwater	8/16	1, 2, 3, 4, 5, 6, 7, 8
Nodivin membrane	<i>cannot</i> dive underwater	8/16	9, 10, 11, 12, 13, 14, 15, 16
Features Related to Nutritional Value			
Calin blood	calcium in blood	2/16	1, 3
Hirn cavity	high in iron	2/16	5, 7
Fibhi arch	high in fiber	2/16	9, 11
Virich tissue	rich in vitamins	2/16	13, 15
Nocalin blood	<i>no</i> calcium in blood	2/16	2, 4
Nohirn cavity	<i>not</i> high in iron	2/16	6, 8
Nofibhi arch	<i>not</i> high in fiber	2/16	10, 12
Novirich tissue	<i>not</i> rich in vitamins	2/16	14, 16
Lofa artery	low in fat	4/16	1, 3, 5, 7
Minhi bones	high in minerals	4/16	9, 11, 13, 15
Nolofa artery	<i>not</i> low in fat	4/16	2, 4, 6, 8
Nominhi bones	<i>not</i> high in minerals	4/16	10, 12, 14, 16
Prolots in gut	lots of protein	8/16	1, 3, 5, 7, 9, 11, 13, 15
Noprolots in gut	<i>not</i> a lot of protein	8/16	2, 4, 6, 8, 10, 12, 14, 16

(Manuscript received August 5, 2005;  
revision accepted for publication May 2, 2006.)