LEARNING PROBABILISTIC RELATIONAL CATEGORIES

Wookyoung Jung
(jung43@cyrus.psych.uiuc.edu)
John E. Hummel
(jehummel@cyrus.psych.uiuc.edu)
Department of Psychology,
603 E. Daniel Street
Champaign, IL 61820, USA

ABSTRACT

Relational concepts play a central role in many cognitive activities, including analogy making. Kittur, Hummel and Holyoak (2004, 2006) showed that people have great difficulty learning relation-based categories with a probabilistic (i.e., family resemblance) structure. In Experiment 1, we investigated three interventions hypothesized to facilitate learning relational categories with a family-resemblance structure: Naming the relevant relations, providing a hint to look for a family resemblance structure, and changing the description of the task from learning categories to choosing the “winning” part of each stimulus, which was predicted to encourage subjects to form an invariant higher-order relation. Only changing the task consistently improved subjects’ ability to learn probabilistic relation-based categories. Experiment 2 investigated the reasons for this “who’s winning” effect. The results suggest that the “who’s winning” task helps both by refocusing subjects’ attention on exemplar parts (rather than the exemplar as a whole) and by invoking schemas for “winning” to help find an invariant higher-order relation that remains constant across all members of a category. Together, the results suggest that relational concept acquisition is greatly facilitated by the discovery of a relational invariant that holds across all members of a category.

INTRODUCTION

One of the most robust findings in the vast literature on category learning is that people are capable of learning categories with a family resemblance structure, in which every member of the category shares some features with every other member, but no single feature is shared by all category members (e.g., Bruner, Goodnow, & Austin, 1956; Kruschke, 1992; Rosch & Mervis, 1975; Shiffrin & Styvers, 1997; Smith & Medin, 1981). The “prototype” effects that result from such learning (such as our ability to learn a category prototype from the exemplars without ever seeing the prototype itself) are so robust that they led Murphy (2002) to quip that any category learning experiment that fails to demonstrate prototype effects is suspect. It is hardly possible to teach a course in cognitive science without talking about prototype effects: They are among the most ubiquitously observed and widely accepted effects in cognitive psychology.

In reviewing the literature on prototype effects, Kittur, Hummel and Holyoak (2004) noticed that all the studies reporting prototype effects had used category structures defined by their members’ features. For example, if the categories to be learned were fictional animals then they might be defined by features such as the shape of the head, the shape of the tail, etc. Similarly, the vast majority of models of category learning and categorization assume that we represent categories and exemplars as lists of features and assign exemplars to categories by comparing their features (see Kittur et al. for a review). As Kittur et al. observed, this reliance on feature-based...
Learning Probabilistic Relational Categories

categories is a limitation inasmuch as many natural concepts and categories are based, not exclusively on features, but also on relations, including both relations between the features of an exemplar (e.g., the seat and back of a chair need to be in a particular spatial relation to serve as a chair) and relations between the exemplar and other objects (e.g., the category conduit is defined by a relation between the conduit and the thing it carries; barrier is defined by the relation between the barrier, the thing to which it blocks access and the thing deprived of that access; even such a basic category as mother is defined by a relation between the mother and her child) (see Gick & Holyoak, 1983; Pirolli & Anderson, 1985; Ross, 1987). The importance of relational categories in human cognition, in combination with their under-representation in one of the largest literatures in cognitive psychology, led Kittur et al. (2004) to pose the following question: Can we observe prototype effects with relational categories? That is, if the categories to be learned are defined by the relations between the exemplars’ features, rather than the literal features themselves, can human subjects learn categories with a family resemblance structure? And if they do, what will the resulting prototypes be like? Kittur et al. never got an answer to the second question because the answer to the first question turned out to be a resounding no: Using a 2 X 2 design crossing category structure (family resemblance, in which no single feature or relation always predicted category membership, vs. deterministic, in which one feature or relation remained invariant across all exemplars of a category) with defining property (exemplar features vs. relations between those features), Kittur et al. found that subjects in the relation/family resemblance condition found category learning much more difficult than subjects in the other conditions (an effect that Kittur, Holyoak & Hummel, 2006, used ideal observer analysis to demonstrate is probably not attributable to the formal difficulty of the task itself); indeed, the majority of the subjects in the relation/family resemblance condition failed to reach criterion even after 600 trials of learning.

Kittur at al. (2004, 2006) interpreted their findings in terms of the LISA model of schema induction (Hummel & Holyoak, 2003). Specifically, they reasoned that if a relational category is represented as a schema, as has been proposed by others (e.g., Barsalou, 1993; Gentner, 1983; Holland, Holyoak, Nisbett, & Thagard, 1986; Keil, 1989; Murphy & Medin, 1985; Ross & Spalding, 1994), and if schemas are learned by a process of intersection discovery, in which a schema is learned from examples by keeping what the examples have in common and discarding details on which they differ (as proposed by Hummel and Holyoak, 2003; see also Doumas, Hummel & Sandhofer, 2008), then learning probabilistic relational categories ought to be extremely difficult because the intersection of the examples is the empty set (i.e., there is no single relation shared by all category members).

In the research reported here, we sought to better understand the difficulty of learning relational categories with a family resemblance structure by investigating circumstances that might make them easier to learn. Experiment 1 tested three not-mutually-exclusive hypotheses about what makes family resemblance relational categories difficult to learn. The results suggest that recasting the category learning task as a task that encourages subjects to discover a higher-order relation that remains invariant across members of a category (namely, as a task that required subjects to learn which part of each exemplar was “winning”) greatly facilitated probabilistic relational category learning. Experiment 2 investigated the reasons for this “who’s winning” effect and suggests that the effect results from a combination of both general

1. Doumas, Hummel, & Sandhofer (2008) predict that intersection discovery is essential, not only to learn entire relational schemas, but also to acquire basic relations, such above and larger-than.
factors, not specific to winning per se, and specific factors.

**EXPERIMENT 1**

Following Kittur et al. (2004), each of our exemplars was composed of two shapes: a square and a circle (Kittur et al. used an octagon rather than a circle, and they placed their stimuli on a background designed to resemble a “computer chip” whereas we did not, but the stimuli are otherwise isomorphic). In each exemplar, one of the two shapes was larger than the other, one was darker, one was in front, and one was above the other. In the prototype of category A (never seen by subjects), the circle was larger, darker, above and in front of the square; in the prototype of category B, the square was larger, darker, above and in front of the circle. In any given exemplar seen by a subject, exactly three of these relations were shared with the prototype of the exemplar’s category and one was shared with the prototype of the opposite category (e.g., an exemplar of A might have the circle larger, darker and above [A-prototype relations] but behind [a B-prototype relation] the square).

The first hypothesis we explored is that people are simply biased toward learning based on features rather than relations. To test this hypothesis, one factor varied whether the instructions to subjects explicitly stated which relations were relevant to category membership. To the extent that the results of Kittur et al. (2004) reflect a bias against using relations for categorization, naming the relations should facilitate category learning.

The second hypothesis we tested was that, rather than being unable to learn relational categories with a family resemblance structure, people are simply biased against assuming that relational categories will have a family resemblance structure. That is, faced with relational categories, perhaps people simply assume that those categories will have some defining (i.e., deterministic; invariant) relation—for example, an essence (see Keil, 1989; Medin & Ortony, 1989)—that is shared by all members of the category, and that this assumption caused Kittur et al.’s subjects to adopt a suboptimal learning strategy. To test this hypothesis, the second factor varied whether a clue was given. In the *clue* condition the instructions explicitly informed subjects that no single property would always work as the basis for categorizing the exemplars. In the *no clue* condition, no such clue was provided. To the extent that subjects are biased against assuming a family resemblance category structure given relational categories, providing this clue should help them to adopt a more appropriate learning strategy, especially when the relations were also named.

Our final hypothesis started with Kittur et al.’s (2004) conclusion: If it is difficult to learn relational categories that have a family resemblance structure, then anything that encourages subjects to discover a property—e.g., a higher-order relation over the first-order relations—that does remain invariant across all members of a category ought to substantially improve relational category learning (since the categories, although probabilistic in the first-order relations, would now be deterministic in the higher-order relation). To test this hypothesis, in the *categorize* condition, subjects were instructed to learn the category of each stimulus, as in Kittur et al. In the *who’s winning* condition, we told subjects they would see displays consisting of a circle and a square, and that in each display “either the circle is winning or the square is winning,” and that their task was to figure out which one was winning. In all other respects, the *who’s winning* task was identical to the *categorize* task: In any stimulus that would be categorized as a member of category A, the circle was “winning,” and in any stimulus that would be categorized as a B, the square was “winning.” The “who’s winning” task could encourage subjects to discover an invariant that holds across members of a “category” by invoking schemas for winning and losing. Such a schema might encourage subjects to predicate a higher-order relation of the form “more winning roles on the circle/square,” which would remain invariant over members...
of a category. If this happens, then even though no (nominally relevant) first-order relation remains invariant over members of a category, the higher-order relation would. If the presence of an invariant is key to the learnability of relational categories (as concluded by Kittur et al., 2004), then subjects in the who’s winning condition might learn faster than those in the categorize condition.

METHOD

Design

The experiment used a 2 (relations named vs. not named) X 2 (clue vs. no clue) X 2 (categorize vs. who’s winning task) between-subjects design.

Procedure

Participants were first given instructions to categorize the stimuli (categorize condition) or decide whether the circle or square was winning (who’s winning task), which either named the relevant relations (relations named) or not (not named) and either provided the “no single property will always work” clue (clue condition) or not (no clue). After the instructions, all conditions were identical. Trials were presented in blocks of 16, with each exemplar presented in a random order once per block. In the categorize condition, subjects were instructed to press the A key if the stimulus belonged to category A and the B key if it belonged to B; in the who’s winning condition, they were instructed to press A if the circle was winning and B if the square was winning (i.e., the stimulus-response mapping was identical across tasks, since in all members of A the circle “wins” and in all members of B the square “wins”). Each exemplar remained on the screen until the participant responded. Responses were followed by presentation of the correct category label. The experiment consisted of 60 blocks (960 trials) and continued until the participant responded correctly on at least fourteen of sixteen trials (87.5% correct) for two consecutive blocks or until all 60 blocks had transpired, whichever came first. At the end of the experiment participants were queried about the strategies they used during the experiment.

Stimuli

Each trial presented a single exemplar consisting of a gray circle and a gray square in the middle of the computer screen. The properties of the exemplars were determined by a family resemblance category structure defined over the relevant first-order relations. The prototypes of the categories were defined as [1,1,1,1] for category A and [0,0,0,0] for B, where [1,1,1,1] represents a circle larger, darker, on top of, and in front of a square and [0,0,0,0] represents a circle smaller, lighter, below and behind a square. Exemplars of each category were made by switching the value of one relation in the prototype (e.g., category A exemplar [1,1,1,0] would have the circle larger, darker, on top of and behind the square). Two variants of each logical structure were constructed by varying the metric properties size and darkness, respecting the categorical relations larger and darker, resulting in eight exemplars per category.

Participants

A total of 154 subjects participated in the study for course credit. Each participant was randomly assigned to one of the eight conditions.

RESULTS AND DISCUSSION

Trials to criterion. Since our primary interest is the rate at which participants learn the categories, we report our data first in terms of trials to criterion. These analyses are biased against our hypotheses in the sense that participants who never learned to criterion were treated as though they reached criterion on the last block. Figure 1 shows the mean trials to criterion by condition. A 2 (relation name vs. no name) X 2 (clue vs. no clue) X 2 (categorize vs. who’s winning) between-
subjects ANOVA revealed a main effect of task \( F(1, 145) = 25.826, \text{MSE} = 2,267,729, p < 0.001 \), reflecting the fact that participants took reliably fewer trials to reach criterion in the who’s winning task \( (M = 211, SD = 261) \) than in the categorize task \( (M = 453, SD = 339) \). No other main effects were statistically reliable. However, here was a reliable interaction between relation naming and clue, indicating that the effect of providing the clue was more pronounced for relation not named than for relation named \( F(1, 145) = 5.98, \text{MSE} = 525,066, p < 0.05 \). Finally, there was a reliable three-way interaction between relation name, clue, and task \( F(1, 145) = 4.10, \text{MSE} = 359,946, p < 0.05 \). As shown in Figure 1, relation naming interacted with the clue differently across the two tasks. With the who’s winning task, the effect of the clue was roughly equivalent to the effect of naming the relations, with each reducing trials to criterion. By contrast, for participants given the categorize task, naming the relations without providing the clue and providing the clue without naming the relations were both beneficial relative to doing neither (although these trends did not reach statistical reliability in our sample); but both naming the relations and providing the clue together did not facilitate category learning, and in fact the trend went in the opposite direction.

Of particular interest is the fact that the condition that gave rise to the worst performance with the categorize task (and overall)—specifically, relation named and clue, with only 50% of participants learning to criterion (and a mean of 623 trials to criterion)—gave rise to the best performance with the who’s winning task (and overall), with 95% of participants learning to criterion (and a mean of 160 trials to criterion). We address the possible reasons for this effect in the Discussion.

**Response Times.** Since participants in the categorize, relations named, and clue condition required so many more trials to reach criterion than participants in the who’s winning, relations named and clue condition, we also analyzed these conditions in terms of participants’ mean response times on individual trials in order to gain insight about the strategies participants in these conditions may have adopted. Response times in the relations named, clue and categorize condition \( (M = 1.69 \text{ s}) \) were reliably shorter than those in the relations named, clue and who’s winning condition \( (M = 3.31 \text{ s}) \) \( t(35) = -4.45, p < 0.001 \). As elaborated in the Discussion, these data suggest that subjects in the former condition were attempting to categorize the stimuli based on their features, whereas those in the latter were attending to the exemplars’

![Figure 1. Mean trials to criterion in the categorize (top) and who’s winning (bottom) conditions.](image)
Learning Probabilistic Relational Categories

relations, including, potentially, higher-order relations.

EXPERIMENT 2

The most striking result in Experiment 1 was the main effect of the who's winning vs. categorize tasks. Accordingly, Experiment 2 sought to further elucidate the reasons for this effect. (The more subtle effects from Experiment 1 are explored in the Discussion.) Specifically, Experiment 2 tested two not-mutually-exclusive hypotheses about how the who’s winning task facilitates learning of probabilistic relational categories: the comparison hypothesis and the specific role of the winning schema itself.

Our first hypothesis was that the who’s winning task facilitates learning simply by encouraging subjects to compare the circle and square in some manner that the category learning task does not. For example, perhaps subjects in the who’s winning condition represented the circle and square as separate objects and doing so facilitated learning by encouraging them to compare them to one another. On this account, any task that encourages subjects to represent the circle and square as separate objects engaged in a relation (like winning/losing) ought to facilitate learning. For example, asking subjects “who’s daxier?” should encourage the same kind of comparison as “who’s winning?” and result in a comparable improvement over “to which category does this example belong?”.

Our second hypothesis was that the winning schema itself may facilitate learning by encouraging subjects to count the number of “winning” roles (i.e., “points”) bound to the circle and the square and to declare whichever part has more winning roles the winner. On this account, the effect of “who’s winning” reflects the operation of the “winning” schema, per se, rather than simply the effect of comparisons encouraged by instructions that suggest the circle and square are separate objects.

Where these hypotheses make divergent predictions is in the role of role alignment in the effect. The instructions refer to the relevant relations by naming one role of each relation: Subjects are told that one shape will be darker, one will be larger, one will be above and one will be in front. Implied, but not stated, is the fact that, therefore, one will be lighter, one smaller, one below and one behind. Perhaps naming darker, larger, above and in front somehow marks them as the “winning” roles, leaving lighter, smaller, below and behind to be the “losing” roles. If so, then to the extent that the effect is due to the involvement of the “winning” schema, per se, then having the roles aligned within categories (i.e., such that the “winning” shape is the one with the most named [i.e., “winning”] roles) ought to lead to faster learning than having the roles misaligned (e.g., such that the “winning” shape that the one that has 3/4 of larger and in front [named, “winning” roles] and lighter and below [unnamed, “losing” roles]). By contrast, to the extent that the effect of “who’s winning” simply reflects the role of comparison, then pole alignment vs. misalignment should make little difference to the rate of learning. A third possibility, of course, is that both hypotheses are correct, in which case we would expect to see facilitatory effects of both comparison (i.e., “who’s daxier?” or “who’s winning?” vs. “what category?”) and, in the case of “who’s winning?” role alignment.

Experiment 2 tested both hypotheses by orthogonally crossing task (categorize vs. who’s daxier vs. who’s winning) with role alignment (aligned vs. misaligned). In all other respects, Experiment 2 was an exact replication of the conditions in Experiment 1 in which participants were informed that what the relevant relations were and that no single relation would work every time.

METHOD

Procedure and Stimuli

Participants were first instructed to categorize the stimuli (categorize condition), decide whether the circle or square was daxier (who’s daxier task) or decide whether the circle or square was winning (who’s winning
task). All instructions named the relevant relations and gave the “no single property will always work” clue.

There were two types of stimuli: In the aligned roles condition, the prototypes were identical to those of Experiment 1. In the misaligned condition, the named roles were mixed across the prototypes of A and B (categorize condition), the “daxier” shape (daxier condition) or the “winning” shape (winning condition). The precise mixing of roles was counter balanced: In one case, the category A prototype/ “daxier” prototype/ “winning” prototype was larger, lighter, below and in front; in the other it was smaller, darker, above and behind. The procedure was identical to that of Experiment 1.

Participants

Participants were 105 undergraduates who participated for course credit. Each participant was randomly assigned to one of the six conditions.

RESULTS AND DISCUSSION

Trials to criterion. The analyses of trials to criterion are conservative as in Experiment 1. The trials to criterion data are shown in Figure 2. A 3 (categorize vs. daxier vs. winning) × 2 (aligned vs. misaligned) between-subjects design ANOVA revealed that a main effect of task \[ F(2, 99) = 11.352, \text{MSE} = 1,158,433, p < 0.001 \]. As in Experiment 1, subjects reached criterion in fewer trials in the who’s winning task \( (M = 381, SD = 360) \) reached criterion faster than those in the misaligned conditions \( (M = 521, SD = 341) \). A reliable difference between aligned \( (M = 206) \) and misaligned \( (M = 468) \) was found only in the winning condition \( \text{t}(36) = -2.534, p < 0.05 \).

![Figure 2. Mean trials to criterion by condition](image)

Response Times. As in Experiment 1, we analyzed response times on individual trials. There was a reliable effect of task \[ F(2,99) = 5.917, \text{MSE} = 9.253, p < 0.01 \] such that RTs in who’s winning \( (M = 2.96, \text{SD} = 1.48) \) were reliably longer than in categorize \( (M = 1.87, \text{SD} = 1.48) \) (by Tukey’s HSD). The difference between aligned and misaligned was not reliable. Experiment 2 thus showed a speed-accuracy tradeoff similar to that observed in Experiment 1.

The results of Experiment 2 are consistent with both our hypothesized explanations of the effect of who’s winning in Experiment 1. The fact that the “who’s daxier” task resulted in faster learning than the categorize task in both the aligned and misaligned roles conditions is consistent with the hypothesis that “who’s winning” (like “who’s daxier”) encourages subjects to compare the circle and square in a way that categorization does not. This hypothesis is further supported by the fact that subjects in the misaligned winning condition performed...
similarly to those in the daxier condition and better than those in the categorize condition. At the same time, the fact that subjects in the aligned winning condition learned faster than those in either the misaligned winning or daxier conditions is consistent with a winning-schema-specific effect. Together, the results of Experiments 1 and 2 suggest that an effective way to help people learn relational categories with a probabilistic structure is to recast the learning task in a form that encourages them to discover a higher-order relation that remains invariant over members of a category.

GENERAL DISCUSSION

Kittur et al. (2004) showed that people find relational categories with a probabilistic structure disproportionately difficult to learn relative to featural categories with a probabilistic structure or relational categories with a deterministic structure. They interpreted this effect in terms of people using schema induction (via intersection discovery) to learn relational categories, an approach that yields useful relational schemas (Doumas et al., 2008; Hummel & Holyoak, 2003) and succeeds with deterministic category structures, but fails catastrophically with probabilistic category structures.

We sought to better understand this phenomenon by investigating conditions under which people might succeed at learning relational categories with a probabilistic structure. Our results showed that recasting category learning as “who’s winning” substantially improved participants’ ability to learn relational categories with a probabilistic structure. Faced with the “who’s winning” task, in Experiment 1, other factors that might sensibly be expected to improve learning—specifically, naming the relevant relations and informing participants that no single relation will work every time—seemed to improve performance (although not all these trends were statistically reliable in our data). Surprisingly, when combined, these factors did not improve the learning of participants charged with the (formally equivalent) task of categorizing the stimuli: Although each factor individually seemed to improve learning of our probabilistic relational categories, when combined they substantially impaired learning.

The reasons for this trend are not entirely clear, but it is consistent with the pattern that would be expected if participants in the relations named, clue and categorize condition were attempting to categorize the exemplars based on their features rather than the relations between them. This conclusion is supported by the fact that response times were fastest in the relations named, clue and who’s winning condition (1.69 s per trial) and slowest in the relations named, clue and who’s winning condition (3.31s per trial). A post-hoc analysis of participants’ end-of-experiment self-reports also supports this conclusion: Participants in the relations named, clue and categorize condition named stimulus features rather than dimensions or relations more often than participants in any of the other conditions (19 times vs. a mean of 8.29 times [SD = 4.39] across the other conditions).

These patterns suggest that participants in Experiment 1’s named, clue and categorize condition may have abandoned the use of the first-order relations as the basis for categorization and, rather than discovering a useful higher-order relation, simply retreated to a strategy based on the exemplars’ features. At the same time, however, it remains unclear why only the participants in this condition would resort to this maladaptive strategy. Perhaps being told what the relevant relations were, in combination with the clue that no single one of them would work every time, had the counterproductive effect of helping these participants know which relations to ignore in their categorizations.

More important for our current purposes is the fact that, as predicted, changing the task from a category learning task to a “who’s winning” task substantially improved our participants’ ability to discover what separated stimuli requiring an “A” response from those requiring a “B” response (a result that obtained in both Experiment 1 and Experiment 2). Importantly, this improvement obtained even
though the formal task—i.e., the specific stimulus-response mappings—were identical across the categorize, winning and daxier conditions. This improvement is consistent with the idea that “who’s winning” encouraged participants to search for an invariant across members of a category. The findings reported here are consistent with Kittur et al.’s (2004) conclusion that learning relational categories is greatly facilitated by the discovery of an abstract invariant that holds true across all members of a category. As such, our data support the idea that relational category learning may entail some form of intersection discovery.

Our findings also suggest that the traditional format of the laboratory category learning task may inhibit the discovery of the invariants necessary for intersection discovery to succeed. Other tasks (such as “who’s winning?”) may be better suited to this purpose. Indeed, the fact that participants in our relations named, clue and categorize condition of Experiment 1 (and the same task in Experiment 2) took the longest of all our participants to reach criterion—and were least likely to reach it—suggests that one of the worst things you can do to a person who is attempting to learn probabilistic relational categories is tell them that they are attempting to learn probabilistic relational categories.

Finally, our results and those of Kittur et al. (2004, 2006) raise the question of whether natural relational concepts tend to have a deterministic or probabilistic structure. Do ad-hoc categories, such as “things to remove from a burning house” and “things to take on a winter camping trip” (Barsalou, 1983) have a relational invariant that holds true across all members of the category? Does our tendency to assume the existence of (invariant) “essences” in biological categories reflect a desire for invariants in relational categories? And do schemas and theories tend to possess relational invariants?

The work of Kittur et al. (2004) suggests that schemas, theories and ad-hoc categories must either contain relational invariants or else be difficult to acquire. The findings presented here suggest that they may not be so difficult to acquire, even if they lack invariants among their first-order relations, provided the conditions under which they are learned promote the discovery of an invariant higher-order relation.

Acknowledgements
This research was supported by AFOSR Grant # FA9550-07-1-0147. We are grateful to David Landy, Eric Taylor, Derek Devnich, Erin Jones and Brian Ross for useful discussions of this work.

REFERENCES


Learning Probabilistic Relational Categories


