MAKING PROBABILISTIC RELATIONAL CATEGORIES LEARNABLE

BY

WOOKYOUNG JUNG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Master's Committee:

Professor John E. Hummel, Chair
Professor Brian H. Ross

ABSTRACT

Kittur, Hummel and Holyoak (2004) showed that people have great difficulty learning relation-based categories with a probabilistic (i.e., family resemblance) structure. In Experiment 1, we investigated interventions hypothesized to facilitate learning family-resemblance relational categories. Changing the description of the task from learning about categories to choosing the "winning" object in each stimulus had the greatest impact on subjects' ability to learn probabilistic relation-based categories. Experiment 2 tested two hypotheses about how the "who's winning" task works. The results are consistent with the hypothesis that the task invokes a "winning" schema that encourages learners to discover a higher-order relation that remains invariant over members of a category. Experiment 3 reinforced and further clarified the nature of this effect. Together, our findings suggest that people learn relational concepts by a process of intersection discovery akin to schema induction, and that any task that encourages people to discover a higher-order relation that remains invariant over members of a category will facilitate the learning of putatively probabilistic relational concepts.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION... 1

CHAPTER 2: EXPERIMENT 1... 11

CHAPTER 3: EXPERIMENT 2... 19

CHAPTER 4: EXPERIMENT 3... 24

CHAPTER 5: GENERAL DISCUSSION... 29

CHAPTER 6: CONCLUSION.. 35

TABLES... 37

FIGURES... 38

REFERENCES.. 43

CHAPTER 1

INTRODUCTION

We use categories to have a better understanding of the world. They allow us to generalize knowledge from one situation to another, to decide which objects in the world are fundamentally the same, and to infer the unseen properties of novel category members (Kittur, Hummel, & Holyoak, 2009). Relational concepts also play a central role in human mental life (Gentner, 1983; Holland, Holyoak, Nisbett, & Thagard; Holyoak & Thagard, 1995; Hummel & Holyoak, 2003). Such concepts specify the relations among two or more things rather than just the literal features of those things. For example, the concept ‘mother’ is defined by the relationship between the mother and her child, not by her status as a female or an adult. Relational categories abound in ordinary language (Gentner, 2005). Asmuth and Gentner (2005) showed that informal ratings of the 100 highest-frequency nouns in the British National Corpus revealed that about half refer to relational concepts, such as ‘barrier’, ‘gate’, ‘conduit’, and ‘bridge’.

The study of human concepts has focused largely (although not exclusive) on category learning, the intuition being that understanding how people acquire concepts (i.e., mental representations of categories) can shed light on the nature of those concepts. However, in spite of their importance, comparatively little research has investigated the acquisition of relational categories. Rather, research on category learning has focused almost exclusively on how people learn categories defined by their exemplars’ features.

One of the most robust findings from this literature is that human concepts have a *family resemblance* structure (e.g., Kruschke, 1992; Kruschke & Johansen, 1999; Rosch & Mervis, 1975; Shiffrin & Styvers, 1997; Smith & Medin, 1981), such that every member of a category tends to share features with other members of the category but there need not be any single defining feature that is shared by all category members. To the extent that relational concepts are qualitatively similar to feature-based concepts, our understanding of concepts can be expected to generalize from the (extensively investigated) case of feature-based categories to the (largely neglected) case of relational categories. But as elaborated shortly, there is reason to believe they are not, casting doubt on our ability to generalize our conclusions from studies using feature-based categories to the case of relational concepts.

We begin with an overview of the literature on feature-based and relational categories. The conclusion of this review is that relational categories, in contrast to feature-based categories, may not be learnable when they have a family-resemblance structure (see also Gentner & Kurtz, 2005; Kittur, Holyoak & Hummel, 2006; Kittur, Hummel & Holyoak, 2004). We then report the results of three experiments investigating the conditions under which people can (and cannot) learn relational categories with a family-resemblance structure. Our findings suggest that the best way to make family-resemblance relational categories learnable is to structure the learning task in a way that endows the to-be-learned relational categories with a higher-order relation that remains invariant over members of a category—i.e., to endow them with a deterministic, rather than a family resemblance, structure.

Feature-based Categories and Prototype Effects

The idea that concepts have a family-resemblance structure is the accepted wisdom in cognitive psychology, but this was not always the case. The “classical” view of concepts—which was dominant since the time of the ancient Greek philosophers—held that concepts specified the necessary and sufficient conditions for category membership (see, e.g., Bruner, Goodnow & Austin, 1956). For example, the concept of a triangle specifies that it is a two-dimensional geometric form consisting of three straight lines that meet at angles summing to 180 degrees. However, Wittgenstein (1953) famously refuted this idea, pointing out that people cannot provide such definitions for many ordinary categories, such as “game” (an argument later empirically supported by Hampton, 1981, 1995).

This failure of the classical view gave rise to the idea that concepts have a family resemblance structure. Instead of a definition, the mental representation of a category was assumed to be either a prototype or a collection of stored exemplars consisting of probabilistic clusters of features (e.g., Rosch & Mervis, 1975; Medin & Schaffer, 1978). This idea motivated an explosion of experiments investigating prototype effects in artificial categories constructed to have a family-resemblance structure. One of the most powerful findings from these experiments is that people are capable of learning categories with a family resemblance structure (e.g., Kruschke, 1992; Kruschke & Johansen, 1999; 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, and the tendency to judge the (unseen) prototype as the “best” member of the category—are so robust that they led Murphy (2002) to quip that any category learning experiment that fails to demonstrate prototype effects is suspect. And it is hardly possible to teach a course on cognitive psychology without talking about prototype effects: They are among the most ubiquitously observed and widely accepted effects in cognitive psychology (see Murphy, 2002).

However, one limitation of these experiments is that they have universally been conducted using categories defined by their exemplars’ features (see Kittur et al., 2004). But as pointed out by numerous researchers, categories defined by simple lists of features are not representative of human concepts (see e.g., Gentner, 1983). Keil (1986, 1989) showed that the surface features of an object could be dramatically changed without changing its classification. For example, people consider a skunk painted and trained to climb trees like a raccoon to still be a skunk, not a raccoon.

Similarly, greater featural overlap does not inevitably result in increased probability of classification as a category member. Malt (1991) demonstrated that people classify liquids with many other substances in them as "water" (e.g., swamp water), whereas they do not classify many liquids that are mostly H_2O as "water" (e.g., juices or sodas). When Rips (1989) asked participants to judge an ambiguous circle that was half way between a quarter and a pizza in size, they judged it to be more similar to a quarter, yet more likely to be a pizza. Gelman and Markman (1986) reported similar results with children, in that children's inductive inferences (about what an animal would feed its babies) were driven more by the animal's category membership than by its perceptual features.

One implication of these studies is that human concepts specify not just the features of categories and their members, but also relations, both between the features of a category (e.g., only if a bird can fly does it also live in trees) and between members of one category and the members of others (e.g., a mother and her child stand in a particular kind of relation to one another). For our current purposes, the crucial difference between a feature and a relation is that a feature is a property of a single entity, whereas the value of a relation cannot be determined without reference to two or more entities. For example, a "feature" of a given circle may be that it is two inches in diameter; however, whether the circle is *larger* or *smaller* than another circle (a relational property) cannot be known without also knowing the size of the other circle (see also Barr & Caplan, 1987).

Relational Categories

There are known differences between featurally-defined and relationally-defined categories. Prior work on relational categories has distinguished between perceptual features and functional features (Bruner, Olver, & Greenfield, 1966; Miller & Johnson-Laird, 1976) and between concrete and abstract features (e.g., Paivio, 1971). Barr and Caplan (1987) explored the distinction between intrinsically represented and extrinsically represented categories. They defined an *intrinsic feature* as one that is true of an entity considered in isolation (such as "has wings" for a bird) and an *extrinsic feature* as one expressing the relationship between two or more entities (e.g., "used to work with" for a hammer). Using a variety of converging measures, they divided their 13 categories into an intrinsically based set (in our terms, feature categories: mammals, birds, flowers, fruit, vegetables and trees) and an extrinsically based set (relational categories: weapons, vehicles, furniture, toys, tools, and sports) with one intermediate category (clothing). Barr and Caplan (1987) found that subjects' responses on a category membership judgment task (e.g., "is an *x* a *y*?") had greater variance (less between-subjects agreement) for relational categories than for featural categories.

Barsalou's (1983, 1985) *ad-hoc* categories, such as "things to take out of the house in case of fire," can also be seen as relational categories. The members of such categories typically lack intrinsic (i.e., featural) similarity. For example, "things to take out of the house in case of fire" includes such diverse exemplars as pets, family photos, checkbooks, laptops, and so forth (see also Estes, Golonka & Jones, 2011, for a discussion of *thematic* relations). In a delightfully circular way, the only "feature" the members of this category have in common is that they are all things to take out of the house in case of fire. In contrast to the members of feature-based categories, which have a graded structure around a central tendency (i.e., the prototype), Barsalou (1985) found that ad-hoc categories show a graded structure around an ideal (properties that optimally promote goal resolution). For example, a food with zero calories for “foods to eat when on a diet”. The centrality of a specific goal suggests that relational category representations may have a relatively sparse, rule-like nature (e.g., see also Kittur et al., 2004; Kittur et al., 2006).

Rehder & Ross (2001) investigated a kind of relational category they termed *abstract coherent categories* and found that such categories can be acquired on the basis of relationships that are orthogonal to the specific attributes of exemplars, as long as the relationships arise in a manner consistent with prior expectations. In their study, they presented three exemplars of the abstract coherent category “morkels”. For example, one morkel “operates on the surface of water, works to absorb spilled oil, coated with spongy material,” while another “operates on land, works to gather harmful solids, has a shovel”. The members of this category lack featural overlap but take their structure from systems of features that support a common abstract relation (i.e., that a morkel’s features work sensibly together to satisfy a goal).

Another kind of relational category is Markman and Stilwell's (2001) *role-governed categories*. Role-governed categories are categories defined by the role an item plays in a more global relational structure. Examples include *private* (the military rank), *mother, mentor, boss*, and so on. Whether someone is a private in the army depends on their relative position in the military hierarchy. Having a single stripe on his or her uniform sleeve may be necessary to recognize a private, but although these features are useful for recognizing whether someone is a private, they are not critical for membership in the category private (e.g., if I were to put a stripe on my sleeve, that would not render me a private) (Markman & Stilwell, 2001). It is in this sense that private is a role-governed category. Although Markman and Stilwell distinguish role-governed categories (which are defined by an exemplar's occupying a particular relational role) from relational categories (which, in their terminology refers only to categories defined by complete relations rather than relational roles, e.g., robbery, betrayal, and slavery) what is important for our current purposes is that both are defined relationally rather than featurally; as such, in our parlance both are relational.

Gentner and her colleagues (Anggoro, Gentner, & Klibanoff, 2005; Gentner, 2005; Gentner & Kurtz, 2005) have also investigated relational categories. They compared traditional *entity* categories with relational categories. According to their definition, entity categories can thought of as first-order partitions of the world (e.g., Gentner, 1982) and relational categories as second-order ways of organizing and linking those first-order partitions (Gentner & Kurtz, 2005). Entity categories such as dog or rose share intrinsic features among the members, often including perceptual commonalities (Gentner, 2005). For example, dogs are characterized by four legs, a tail, fur, a bark, and so on. By contrast, for members that enter into relational categories, there exists no obvious intrinsic similarities among members. They are connected by a sparse relational structure. For example, landlords are defined by the rental relations between a tenant and a landlord, not by perceptual features that landlords have in common.

Like Markman and Stilwell (2001), Gentner and Kurtz (2005) also noted that relational categories can be divided into *relational role* categories (Markman & Stilwell's *role-goverend categories*) and *relational schema* categories (Markman & Stilwell's *relational categories*); but like us, they emphasize the fact that both are inherently relational in nature. Gentner (1982) compared entity categories with relational categories using an exemplar generation task and found that not only were people less fluent at generating members of relational categories, they were also less able to guess the category given the exemplars. This result, along with the findings of Barr and Caplan (1987), suggests that, although relational categories play a central role in our mental lives, they are generally more abstract and more variable across individuals. Thus they're more difficult to reason about than featural categories.

Thematic relations can be seen as yet another kind of relational category. A thematic relation is any temporal, spatial, causal or functional relation between things that perform complementary roles in the same scenario (Estes, Golonka & Jones, 2011, Golonka & Estes, 2009; Lin & Murphy, 2001; Wisniewski & Bassok, 1999). Examples of thematic relations include the relation between bees and honey and the relation between coffee and cream. Thematic relations are *external* in that they occur between multiple objects, concepts, people or events and *complementary* in the sense that the arguments of a thematic relation fill complementary roles of the relation (e.g., bees do the making and the honey gets made; Estes et al., 2011). In this way, the arguments of a thematic relation differ from members of an ad-hoc category: There is no sense in which the members of an ad-hoc category are bound to complementary roles. The sense in which thematic relations form (typically very small) categories is that the arguments of a thematic relation can be viewed as the "members" of the category.

What Must Be Explained?

In reviewing the literature on prototype effects, Kittur et al. (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 categories is a limitation inasmuch as many natural concepts and categories are based exclusively on relations.

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 et al., 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 training.

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*, which keeps what the examples have in common and discards details on which they differ (as proposed by Hummel & 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 the results suggest that the effect results from a combination of both general factors, not specific to winning per se, and winning−specific factors. Experiment 3 replicated and extended Experiment 2. The results of all three experiments suggest that the best way to make a probabilistic relational category learnable is to structure the learning task in such a way as to render the category structure effectively deterministic.

CHAPTER 2

EXPERIMENT 1

Experiment 1 served as an initial investigation of the factors that might make probabilistic relational categories learnable and was motivated by two broad hypotheses. The first (which subsumes two more specific hypotheses as elaborated shortly) was that subjects may simply be biased against learning probabilistic relational categories—or that learning such categories may simply be more difficult (although not qualitatively different) than learning featural categories. As a test of this hypothesis, we manipulated how explicit the instructions were about the fact that the categories were relational and how explicit they were about the fact they were probabilistic. To the extent that subjects are biased against expecting categories to be relational or probabilistic, instructions directing them to expect probabilistic relational categories may improve learning relative to a neural instruction baseline. The second broad hypothesis was that, if relational categories are learned by a process akin to schema induction, then changing the task from a category learning task to a task that can help subjects find a (higher-order) relation that remains invariant over category members should improve performance relative to a category learning baseline.

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 (Figure 1). 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 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 and thus pointing out that relations, rather than features are relevant, 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 Gelman, 2000; 2004; 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 subjects were informed about the probabilistic category structure. 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 correctly 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 the key to the learnability of relational categories (as concluded by Kittur et al., 2004), then subjects in the w*ho's winning* condition might learn faster than those in the *categorize* condition.

Method

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

Materials. Each trial presented a single exemplar consisting of a gray circle and a gray square in the middle of the computer screen (although both figures were gray, they could be darker or lighter shades of gray). 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.

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, the procedure was identical across all conditions. 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 or 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 or winning shape. 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 had used during the experiment.

Results

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 had reached criterion on the last block. Figure 2 shows the mean trials to criterion by condition. A 2 (*relations named* vs. *not named*) × 2 (*clue* vs. *no clue*) × 2 (*categorize* vs. *who's winning*) between-subjects ANOVA revealed a main effect of task [$F(1, 145) = 25.826$, $MSE = 2{,}267{,}729$, $p < 0.001$], reflecting the fact that participants took reliably fewer trials to reach criterion in the *who's winning* condition ($M = 211$, $SD = 261$) than in the *categorize* condition ($M = 453$, $SD = 339$). No other main effects were statistically reliable. However, there was a reliable interaction between *relations named* and *clue*, indicating that the effect of providing the clue was more pronounced for relation *not named* than for *relations named* [$F(1, 145) = 5.98$, $MSE = 525{,}066$, $p < 0.05$]. Finally, there was a reliable three-way interaction between *relations named*, *clue*, and *task* [$F(1, 145) = 4.10$, $MSE = 359{,}946$, $p < 0.05$]. As shown in Figure 2, 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, *relations 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.

Survival function. We also analyzed how many participants reached criterion by the end of each block. The resulting survival functions for the *who's winning* and *categorize* conditions are shown in Figure 3. (A participant's having "survived" on a given block means they had not yet reached criterion by that block. Participants who survived to block 60 never reached criterion.) As shown in Figure 3, a higher proportion of participants reached criterion in the *who's winning* condition than in the *categorize* condition, and they did so much faster.

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 *categorize, relations named*, and *clue* condition (M = 1.69 s) were reliably shorter than those in the *who's winning, relations named*, and *clue* condition (M = 3.31 s) [$t(35) = -4.45$, $p < 0.001$]. As elaborated in the Discussion, these data suggest that participants in the former condition were attempting to categorize the stimuli based on their features, whereas those in the latter were attending to the exemplars' relations, including, potentially, higher-order relations.

Discussion

The results of Experiment 1 showed that recasting category learning as a "who's winning" task substantially improved participants' ability to learn probabilistic relational categories. For subjects given the "who's winning" task, 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 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 *categorize, relations named*, and *clue* condition were attempting to categorize the exemplars based on their features rather than the relations between the circle and square. This conclusion is supported by the fact that response times were fastest in the *categorize, relations named*, and *clue* condition (1.69 s per trial) and slowest in the *who's winning, relations named*, and *clue* 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 *categorize, relations named*, and *clue* 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. Experiment 2 investigated the origin of the winning effect by comparing it to related tasks. Our second study also provided an opportunity to replicate the basic findings of Experiment 1.

CHAPTER 3

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. 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 might facilitate 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 a schema for what "winning" consists of 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 potential role of *role alignment* in the effect. The instructions refer to the relevant relations by naming one role of each: 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 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 role 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 of what the relevant relations were and that no single relation would work every time.

Method

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

Materials. 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 /"daxier"/ "winning" prototype was *larger*, *lighter*, *below* and *in front*; in the other it was *smaller*, *darker*, *above* and *behind* (where *larger*, *darker*, *above* and *in front* were named in the instructions and were thus presumably the "winning" roles).

Design. The experiment used a 3 (*categorize* vs. *who's daxier* vs. *who's winning task*) X 2 (*aligned* vs. *misaligned roles*) between-subjects design.

Procedure. The procedure was identical to that of Experiment 1. Participants were first instructed to categorize the stimuli (*categorize* condition), decide whether the circle or square was daxier (*who's daxier* condition) or decide whether the circle or square was winning (*who's winning* condition). All instructions named the relevant relations and gave the "no single property will always work" clue.

Results

Trials to criterion. The analyses of trials to criterion are conservative in the same sense as in Experiment 1. The trials to criterion data are shown in Figure 4. A 3 (*categorize* vs. *daxier* vs. *winning*) × 2 (*aligned* vs. *misaligned*) between-subjects design ANOVA revealed a main effect of task [$F(2, 99) = 11.352$, $MSE = 1{,}158{,}433$, $p < 0.001$]. As in Experiment 1, participants reached criterion in fewer trials in the *who's winning* task ($M = 330$, $SD = 342$) than in the *categorize* task (*M* = 699, *SD* = 317) (by Tukey's HSD, $p < 0.01$). Participants given the *who's daxier* task (*M* = 492, *SD* = 318) took reliably fewer trials to reach criterion than those in the *categorize task* (*M* = 699, *SD* = 317) (by Tukey's HSD, $p < 0.05$). Participants given the *who's winning* task took fewer trials to reach criterion those given *who's daxier* (by Tukey's HSD, $p < 0.05$). There was also a reliable main effect of alignment [$F(1, 99) = 4.701$, *MSE* = 479,678, $p < 0.05$]. As expected, participants in the aligned conditions (*M* = 381, *SD* = 360) reached criterion faster than those in the misaligned conditions (*M* = 521, *SD* = 341). This difference between aligned (*M* = 206, *SD* = 279) and misaligned (*M* = 468, *SD* = 359) was reliable only in the *who's winning* condition [$t(36) = -2.534$, $p < 0.05$].

Response times. As in Experiment 1, we analyzed response times on individual trials. There was a reliable effect of task [$F(2,99) = 9.296$, *MSE* = 13.588, $p < 0.001$] such that RTs in *who's winning* (*M* = 2.96, *SD* = 1.48) were reliably longer than in *categorize* (*M* = 1.66, *SD* = 0.75) (by Tukey's HSD, $p < 0.001$) and RTs in *who's daxier* (*M* = 2.55, *SD* = 1.18) were reliably longer than in *categorize* (by Tukey's HSD, $p < 0.05$). The main effect of *aligned (M* = 2.56, *SD* =1.45*)* and *misaligned (M* = 2.36, *SD* = 1.14*)* was not reliable ($F(1, 99) = 0.402$, *MSE* = 0.588, $p = 0.527$). Experiment 2 thus showed a speed-accuracy tradeoff similar to that observed in Experiment 1.

Discussion

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 *who's daxier* resulted in faster learning than *categorize* in both the aligned and misaligned 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.

CHAPTER 4

EXPERIMENT 3

The results of Experiments 1 and 2 suggest that finding an invariant higher-order relation is extremely helpful to learning relational categories with a probabilistic structure and that the *who's winning* task also plays an important role in finding such an invariant with our stimuli. Experiment 2 also demonstrated that simply having a task that encourages subjects to think of the circle and square as separate objects (the *who's daxier* task) is not, by itself, sufficient to achieve the same degree of facilitation as enjoyed by subjects given the *who's winning* task. By itself, however, the difference between *who's winning* and *who's daxier* is not sufficient to conclude that something like a "winning schema" is responsible for subjects' superior performance in the *who's winning* condition than in the *categorize* condition.

Specifically, there are at least two additional differences between *who's winning* and *who's daxier* that could account for the superior performance in the former condition: First, the question *who's winning*? is simply more meaningful than the question *who's daxier*?, so it is at least logically possible that this difference in meaningfulness somehow led to better performance in the *who's winning* condition. And second, asking "who's winning?" implies that whoever is not winning is losing. That is, the two roles of the winning/losing relation have opposite valence. Perhaps it is something about relational roles with opposite valence, rather than winning per se, that encourages subjects to invoke a schema that facilitates the discovery of an invariant higher-order relation with our stimuli.

Experiment 3 was designed to tease apart these possibilities. Subjects performed one of five different tasks: The *categorize*, *who's winning* and *who's daxier* tasks were the same as in Experiment 2. In addition, one group of subjects was asked to learn "which one would Britney

Spears like?" We chose this task because, like *who's winning* and *who's daxier*, it encourages subjects to think of the circle and square as separate objects. And like *who's winning*, but unlike *who's daxier*, its roles have opposite valence (presumably it is "good" to be liked by Britney and bad not to be liked by her) and it has meaning. A fifth group of subjects were asked to learn "which one comes from Nebraska". This task shares the comparative property of *winning*, *daxier* and *Britney* and it has semantic content, like *winning* and *Britney*, but presumably lacks strong differences in valence across its roles (i.e., it is presumably neither particularly good nor particularly bad to be from Nebraska). These properties of our five tasks are summarized in Table 1.

To the extent that simply having semantic content is sufficient to account for the difference between *who's winning* and *who's daxier*, performance in the *Britney* and *Nebraska* conditions should resemble performance in *who's winning* and be better than performance in *who's daxier*. To the extent that having asymmetrical valence across relational roles is sufficient, performance in *Britney* should resemble that in *who's winning* but performance in *Nebraska* should resemble performance in *who's daxier*.

In addition, as in Experiment 2, we crossed the five learning conditions orthogonally with role alignment vs. misalignment as an additional check on their similarity to *who's winning* or *who's daxier*. To the extent that Britney vs. Nebraska are like *who's winning* vs. *who's daxier* they should show the same patterns of sensitivity vs. insensitivity to role alignment.

Method

Participants. 191 undergraduates participated in the experiment to fulfill a course requirement. Each participant was randomly assigned to one of the 10 conditions.

Materials. The same stimuli used in Experiments 1 and 2 were in Experiment 3. The prototypes and category structures were identical to those of Experiment 2.

Design. The experiment used a 5 (c*ategorize* vs. *who's daxier* vs. *which one comes from Nebraska vs. which one would Britney Spears like* vs. *who's winning*) X 2 (*aligned* vs. *misaligned roles*) between-subjects design.

Procedure. Aside from the instructions in the *Britney* and *Nebraska* conditions, the procedure was identical to that of Experiment 2.

Results

Trials to criterion. We conducted A 5 tasks (c*ategorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) X 2 alignment (*aligned* vs. *misaligned roles*) ANOVA, but Levene's test (Levene, 1960) for equality of variance revealed a significant difference in variance across five groups in the trials to criterion ($p < 0.001$). The frequency results of trials to criterion across groups were positively skewed. Thus, the data were log transformed to normalize the skewed distributions. After this transformation, a 5 tasks (c*ategorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) X 2 alignments (*aligned* vs. *misaligned roles*) between-subjects design ANOVA revealed that Levene's test was not significant ($p = 0.588$). As shown in Figure 5, there was a main effect of the task [$F(4, 181) = 9.811$, $MSE = 7.108$, $p < 0.001$]. Since our main interest is in how the task itself affects the learning of probabilistic relational categories, we first report the data from the aligned condition. A 5 tasks (c*ategorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) between-subjects ANOVA revealed a reliable effect of the task [$F(4, 106) = 11.150$, $MSE = 7.525$, $p < 0.001$]. As in the previous experiments, participants in *who's winning* (M = 127, SD = 127) reached criterion faster than those in *who's daxier* (M = 276, SD = 241) (p < 0.01) as well as those in *categorize* (M = 523, SD = 341) ($p < 0.001$) (by Tukey's HSD).

Participants in *daxier* took reliably fewer trials to reach criterion than those in *categorize* (by Tukey's HSD, $p < 0.05$). Participants in *Britney* ($M = 186$, $SD = 155$) showed performance equivalent to those in winning. There was no reliable difference between *Britney likes* and *winning* (by Tukey's HSD, $p = 0.381$). Participants in *Britney* reached criterion reliably faster than those in *categorize* (by Tukey's HSD, $p < 0.001$). The number of trials to reach criterion in *Nebraska* ($M = 250$, $SD = 218$) was reliably less than in *categorize (*by Tukey's HSD, $p < 0.05$). There was also a main effect of alignment [$F(1, 181) = 26.769$, $MSE = 19.394$, $p < 0.001$]. The *winning* task [$t(38) = 3.717$, $p < 0.01$], the *Britney* task [$t(33) = 3.036$, $p < 0.01$], and the *Nebraska* task [$t(35) = 3.006$, $p < 0.01$], all of which have semantic content, showed reliable differences between the aligned and misaligned conditions. Whereas there was no such effect, in *categorize* ($t(37) = 1.308$, $p = .199$] or *daxier* [$t(38) = 0.388$, $p = 0.7$] which lack semantic content. There was no reliable interaction between task and alignment [$F(4, 181) = 2.208$, $MSE = 1.6$, $p = 0.07$].

Response times. Levene's test revealed a significant difference between in variance across five groups in response times ($p < 0.05$). The data were log transformed to normalize the skewed distribution. After the transformation, a 5 tasks (c*ategorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) X 2 alignments (*aligned* vs. *misaligned roles*) between-subjects design ANOVA revealed that Levene's test was not significant ($p = 0.178$). We report response times in the *aligned* condition since we were mainly interested in how the different tasks affect category learning. A 5 tasks (c*ategorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) between-subject design ANOVA revealed a reliable effect of task [$F(4,106) = 6.177$, $MSE = 0.940$, $p < 0.001$]. Participants in *winning* ($M = 3.43$, $SD = 1.73$) (by Tukey's HSD, $p < 0.001$) and *daxier* ($M = 2.89$, $SD = 1.29$) (by Tukey's HSD, $p < 0.05$) took reliably longer to respond than those in *categorize* (*M* = 1.88, *SD* = 0.66). RTs in *Britney* (*M* = 2.58, *SD* = 0.79) were marginally longer than RTs in *categorize* (*p* = 0.06*)* and RTs in *Nebraska* (*M* = 2.69, *SD* = 1.34) were also marginally longer than RTs in *categorize* (*p* = 0.06) (by Tukey's HSD). As in Experiments 1 and 2, a speed-tradeoff was found in Experiment 3.

Discussion

Experiment 3 explored why *who's winning* promotes faster learning of probabilistic relational categories than *who's daxier*. The *who's winning* task is semantically rich, has roles of opposite valence and encourages subjects to consider the circle and square as separate objects. To examine how each of these factors contributes to the acquisition of an invariant higher-order relation, we added two new tasks, *which one would Britney Spears like*, which was assumed to have all three of these elements and thus to be equivalent to *who's winning*, and *which one comes from Nebraska* which was assumed to have the first and third, but without opposite valence. In contrast to the *Nebraska* task, the *daxie*r task has no semantic content. Consistent with our assumptions, it was revealed that *Britney* is similar to *winning*, and *Nebraska* is similar to *daxier*. That is, like *winning* and *Britney*, the tasks that treat the circle and square as separate objects, have roles with opposite valence, and have semantic content may provide optimal conditions for discovering a higher-order invariant and thus facilitate learning. Missing of any of these elements, however, seems to make category learning reliably worse.

CHAPTER 5

GENERAL DISCUSSION

Kittur and colleagues (2004) showed that learning relational categories with a probabilistic (family resemblance) structure is extremely difficult. They interpreted this result as indicating that relational category structures invoke the machinery of schema induction by intersection discovery (Hummel & Holyoak, 2003), a learning algorithm that works well for deterministic structures but fails catastrophically with probabilistic category structures, in which no single feature or relation remains invariant across all exemplars of a category. The current study further tested this *intersection discovery* hypothesis by exploring the conditions that might render probabilistic relational category structures learnable. Specifically, the intersection discovery hypothesis predicts that any task that leads learners to discover a (for example, higher-order) property or relation that remains invariant over exemplars of an otherwise probabilistic relational category ought to render that category learnable.

Experiment 1 showed that replacing the category-learning task with the completely isomorphic task of learning which of two parts of an exemplar is "winning" renders the probabilistic relational category structures easily learnable: Although people had great difficulty learning whether a given circle-square pair belonged to category A or category B, they had no difficulty learning whether the circle or the square was "winning", even though the stimulus-response mappings were identical across the two tasks. And although naming the relevant relations and providing the clue (that no single relation would be a reliable indicator of the correct response every time) both facilitated learning in the *who's winning* condition, they individually facilitated—but, together, interfered with—learning in the *categorize* condition. Indeed, learning was numerically (although not reliably) slowest in the *relations named*, *clue* and *categorize* condition: It would appear that the worst thing one can do to a person who is trying to learn probabilistic relational categories is inform them that they are trying to learn probabilistic relational categories.

Experiment 2 demonstrated that this effect is not simply due to subjects' being encouraged to compare the circle and square as separate objects in the "winning" task: Asking subjects "which is daxier" improved learning relative to the category-learning task but did not bring it up to the level of performance in the "winning" task. Finally, Experiment 3 systematically explored the properties of the "winning" task by comparing it to a variety of related learning tasks. The results suggest that what is crucial about the "winning" task is that it encourages learners to compare the circle and square with the goal of assigning each to one role of a relation whose roles have unequal valence.

These results replicate and extend those of Kittur et al. (2004, 2006), providing further evidence that the task of learning a category defined by the relations between things (rather than just the features of the things themselves) invokes a process of schema induction by intersection discovery in the mind of the learner. Although this process works well with categories whose members all share one or more invariant relations, it fails catastrophically with categories that lack such an invariant.

The Frailty of Probabilistic Relational Concepts

Why should relational category learning be so fragile in the face of probabilistic (family resemblance) category structures whereas featural category learning is not? Although the data presented here cannot provide a definitive answer to this question, it is tempting to speculate that this distinction between the learning of feature- vs. relation-based categories reflects differences in the learning algorithms brought to bear on the two tasks.

Feature-based category structures are easy to learn by simple association: It is only necessary to tabulate, explicitly or implicitly, the co-occurrence statistics of features and category labels. And as long as the majority of features in exemplar favor one category label over another, it is possible to assign the correct category label to the stimulus. (To wit: the category structures used in most feature-based studies of category learning are learnable by an one- or at most two-layer associative learner, such a connectionist model trained by error back-propagation; see, e.g., Krushke, 1992.) Deterministic featural categories might be easier to learn than probabilistic ones, but only because a larger majority of the features “point in the direction” of one category label or another, or because one feature (i.e., the deterministic one) “points” more strongly. Deterministic category structures are thus not qualitatively different than probabilistic ones; they are simply different points on the same associative learning continuum.

Although associative learning is adequate for acquiring non-relational (e.g., feature-based) concepts, it is inadequate for acquiring relational (equivalently, *symbolic*; see Hummel, 2010; Hummel & Holyoak, 2003) concepts (see Chomsky, 1959; Deacon, 1997; Doumas et al., 2008; Hummel & Holyoak, 2003). The reason, in brief, is that the meaning of a relational concept like *larger-than* is simply nowhere to be found in the co-occurrence statistics of the objects or features engaged in that relation (see, e.g., Doumas et al., 2008; Gentner, 1983): Almost any given object is both *larger-than* and *smaller-than* a veritable infinity of other objects (a point and the universe as a whole, respectively, being the only notable exceptions). This observation has led some researchers to hypothesize that relational structures such as schemas (Hummel & Holyoak, 2003) and even individual relations (such as *larger-than*; Doumas et al., 2008) are learned, at least in part, by a processes of structured intersection discovery.

The basic idea is that by comparing early, proto-relational ideas (e.g., examples of specific objects in specific, as-yet-discovered relations) to one another *structurally* (i.e., by a process, such as analogical mapping, that highlights the relational correspondences between their parts), intersection discovery serves to highlight what they have in common and de-emphasize the details on which they differ (e.g., comparing an apple to a toy fire truck would highlight properties such as *red* and *shiny* and deemphasize properties unique to one object or the other). Performed iteratively, this operation can result in the discovery of abstract schemas and rules (Hummel & Holyoak, 2003) and even basic relational concepts, such as *larger-than* and *chases* (Doumas et al., 2008). (This proposal is less chicken-and-egg like than it may first appear based on by the very brief summary given here; see Doumas et al., 2008, for the details.)

Hummel and Holyoak (2003) showed that intersection discovery provides a good account of schema induction, and Doumas et al. (2008) demonstrated that it provides an excellent account of the acquisition of relational concepts (in both development and adulthood). But as noted previously, it fails catastrophically with concepts that do not have a deterministic structure.

One potential explanation of the data presented here, and by Kittur and colleagues (2004, 2006), is that both (feature-based) associative learning and (relational) schema induction are engaged during all cases of concept acquisition (see Kittur, et al, 2006; Ashby, Paul & Maddox, 2011, for similar proposals and supporting evidence). Associative learning succeeds in acquiring feature-based categories, perhaps rendering the results of schema induction irrelevant or at least redundant. In response to deterministic relational categories, associative learning fails, but schema induction succeeds, resulting in relational concepts (schemas and/or relational predicates). But faced with probabilistic relational categories, both associative learning and schema induction fail, leaving the learner with little or no basis for categorizing new exemplars.

Psychological Essentialism

The need for an invariant in relational categories is reminiscent of *psychological essentialism*—the belief that (especially biological) categories have an unseen, internal essence that defines them and that gives rise to their visible features (Gelman, 2004; Medin & Ortony, 1989). For example, the visible differences between men and women are assumed to reflect unseen, but fundamental, differences in their makeup (what we now know to be their genes). And when a caterpillar changes into a butterfly, we believe it to be the same animal because, although the visible features have changed considerably, its unseen essence remains the same.

Analogously, we found that the key to making putatively probabilistic relational categories learnable is encouraging the discovery an invariant higher-order relation. Like an "essence", this higher-order invariant is not observable in any single first-order relation. But even as the first order relations change (much as the caterpillar changes into the butterfly), so long as the higher-order invariant—the essence—is preserved, the exemplar remains a member of the category.

Relational Language

Our findings are also broadly consistent with previous findings that relational language —a label that "points to" a relation directly—plays an important role in the acquisition of relational concepts (Casasola, 2005; Gentner, 2003; Gentner, Anggoro, & Kalibanoff, 2011; Gentner, & Christie, 2010; Gentner & Namy, 1999, 2004, 2006; Loewenstein & Gentner, 2005). For example, Gentner and her colleagues (2011) studied children's ability to learn relational concepts (e.g., the *cuts* relation between a knife and watermelon) and found that labeling the relation (e.g., as "dax") facilitated 4- and 5-year-olds' ability to generalize those relations to new arguments (e.g., scissors and paper). The findings showed that the use of relational language encouraged the children to predicate the relational concepts. It is possible that our "who's winning" and "who's daxier" tasks similarly facilitate the acquisition of our relational categories by somehow "pointing to" the crucial relations themselves.

Thematic Relations

Thematic relations (e.g., the relation between bees and honey, or between houses and home owners) are relations between separate objects, rather than relations between the components of a single object. Indeed, it seems intuitive that most natural relational categories are categories defined by relations between the categorized object and other objects, not by relations among the parts of a single object. (But see Biederman, 1987, and Hummel & Biederman, 1992, for arguments that natural object categories also depends on the internal relations between an object's parts.)

These considerations suggest an alternative, or perhaps auxiliary, account of the "who's winning" effect reported here: It is at least possible that the "categorize" task encourages subjects to view the relations as relations among the parts of a single object and the "who's winning" task encourages them to view the critical relations as relations between separate objects.

CHAPTER 6

CONCLUSION

Three experiments showed that recasting category learning as a "who's winning" task considerably improved participants' ability to learn relational categories with a family resemblance structure. Our findings also suggest that the traditional categorize-with-feedback laboratory category learning task may somehow inhibit, or at least fail to promote, the discovery of the higher-order invariants necessary for intersection discovery to succeed with exemplars defined in terms of probabilistic first-order relations. As such, it appears that probabilistic relational categories may be more learnable if one does not realize one is engaged in category learning. Indeed, participants in our relations named, clue and categorize condition of Experiment 1 took the longest of all our participants to reach criterion—and were least likely to reach it—suggesting 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, it is worth noting that our results, like those of Kittur et al. (2004, 2006), raise the question of whether natural relational concepts and categories tend to have deterministic or probabilistic structures. Ad-hoc categories, such as "things to remove from a burning house", "ways to escape the mob" and "things to take on a winter camping trip" (Barsalou, 1983) certainly have a relational invariant that holds true across all members of the category (i.e., category membership). Does our tendency to assume the existence of (invariant) "essences" in biological categories reflect a tendency to assume relational categories possess invariants? And do schemas and theories tend to possess relational invariants? For example, is there a relational core that all members of the category "mother" have in common? Although at first it is tempting to say yes, the differences between birth mothers and adoptive mothers, and between loving mothers and abusive mothers, suggest that the answer might be no. If the answer is no, then do people have difficulty acquiring an all-encompassing schema for the concept “mother”? Or do we simply have multiple “mother” schemas? 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.

TABLES

Table 1

Comparison of three main factors of all conditions in Experiment 3

	Categorize	Who's daxier	Which one comes from Nebraska?	Which one would Britney Spears like?	Who's winning
Treats circle and square as separate objects	No	Yes	Yes	YES	YES
Unequal valence across two relational roles	No	No	No	YES	YES
Has semantic content	No	No	Yes	YES	YES

FIGURES

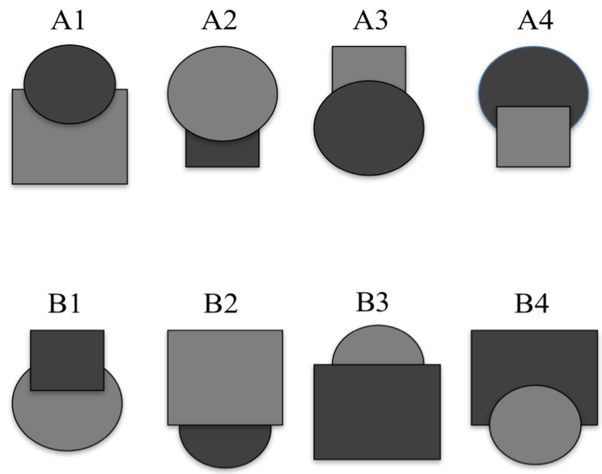

Figure 1. Exemplars of category A and B used in Experiment 1-3. Exemplars of each category were made by switching the value of one relation in the prototype.

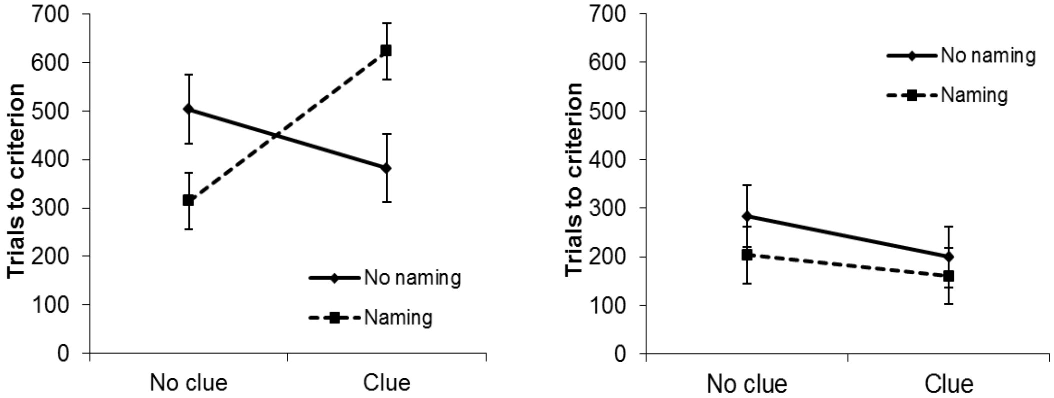

Figure 2. Mean trials to criterion in the categorize (left) and who's winning (right) conditions in Experiment 1. Error bars represent standard errors.

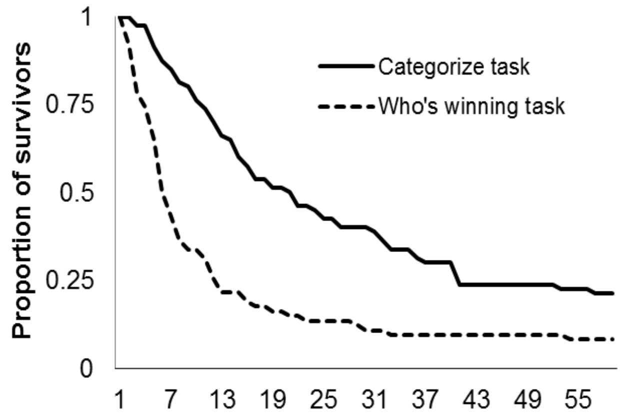

Figure 3. Proportion of survivors per block for each task in Experiment 1. The graph represents how many participants reached criterion by the end of each block.

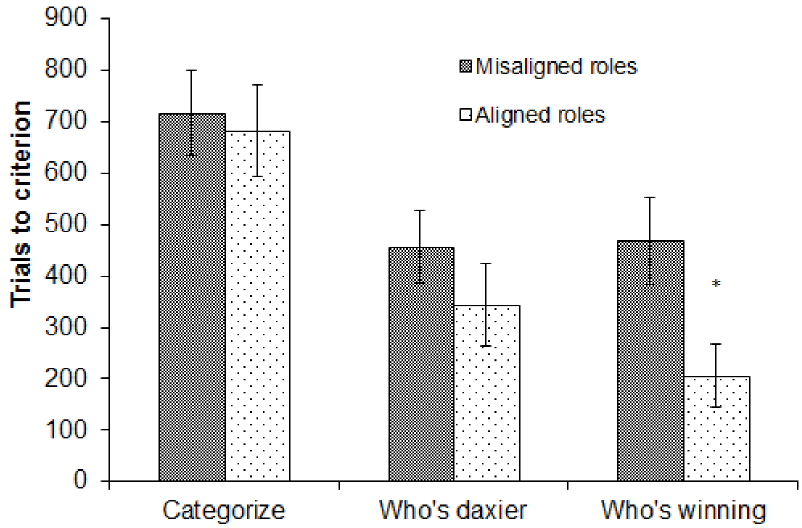

Figure 4. Mean trials to criterion by condition in Experiment 2. Error bars represent standard errors.

*$p < 0.05$

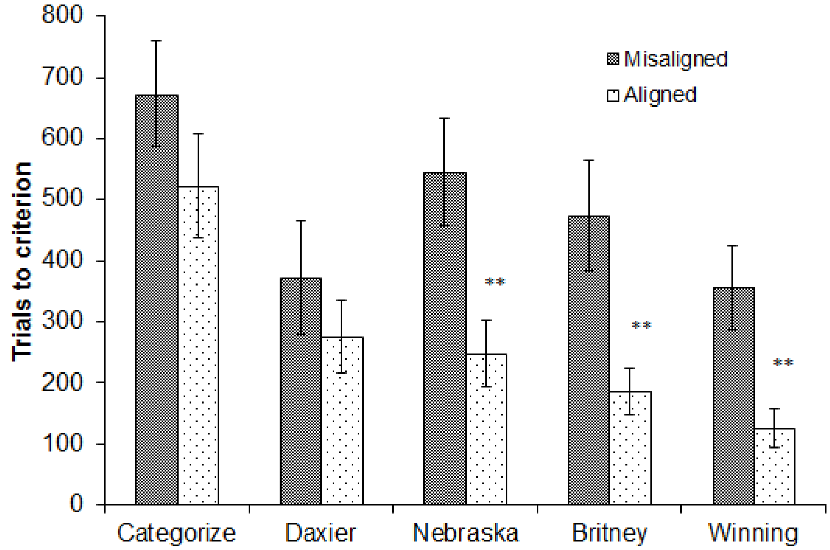

Figure 5. Mean trials to criterion by condition in Experiment 3. Error bars represent standard errors.

**$p < 0.01$

REFERENCES

Anggoro, F., Gentner, D., & Klibanoff, R. (2005). How to go from nest to home: Children's learning of relational categories. *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*, 133-138.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A.J. Wills (Eds.), *Formal approaches in categorization* (pp. 65-87). New York: Cambridge University Press.

Asmuth, J. & Gentner, D. (2005). Context sensitivity of relational nouns. *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*, 163-168.

Barr, R.A., & Caplan, L.J. (1987). Category representations and their implications for category structure. *Memory & Cognition, 15*, 397–418.

Barsalou, L.W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211-227.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Barsalou, L.W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C. Collins, S.E. Gathercole, & M.A. Conway (Eds.), *Theories of memory* (pp. 29-101). London: Lawrence Erlbaum Associates.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding, *Psychological Review, 94*, 115-147.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. Oxford, England: John Wiley and Sons.

Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in cognitive growth.* New York: Wiley.

Casasola, M. (2005). Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology, 41*, 183–191.

Chomsky, N. (1959). Verbal behavior by B. F. Skinner, *Language, 35*, 26-58.

Deacon, T. W. (1997). The symbolic species: The co-evolution of language and the brain. New York: W. Norton and Company.

Doumas. L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1-43.

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: The apprehension and consequences of thematic relations (pp. 249-294). In B. Ross (Ed.), *Psychology of Learning and Motivation, Vol. 54.* Burlington: Academic Press.

Gelman, S. A. (2000). The role of essentialism in children's concepts. In H. W. Reese (Ed.), *Advances in child development and behavior, Vol. 27* (pp. 55-98). San Diego: Academic Press.

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

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183-209.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), Language development: Vol. 2. Language, thought and culture (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155-170.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp.195-235). Cambridge, MA: MIT Press.

Gentner, D. (2005). The development of relational category knowledge. In L. Gershkoff-Stowe & D. H. Rakison, (Eds.), *Building object categories in developmental time*. (pp. 245-275). Hillsdale, NJ: Erlbaum.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure-mapping and relational language support children's learning of relational categories. *Child Development*, *82,* 1173-1188.

Gentner, D., & Christie. S. (2010). Mutual bootstrapping between language and analogical processing. *Language and cognition*, *2*, 261-283.

Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), *Categorization inside and outside the lab*. (pp. 151-175). Washington, DC: APA.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487-513.

Gentner, D., & Namy, L. L. (2004). The role of comparison in children's early word learning. In D. G. Hall & S. R. Waxman (Eds.), Weaving a lexicon (pp. 533-568). Cambridge: MIT Press.

Gentner, D. & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15*, 297-301.

Golonka, S. & Estes, Z. (2009). Thematic relations affect similarity via commonalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1454-1464.

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, *9*, 149-156.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language*, *34*, 686-708.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of Inference, Learning, and Discovery*. Cambridge, MA, US: The MIT Press.

Holyoak, K. J., & Thagard, P. (1995). Mental leaps: Analogy in creative thought. Cambridge, MA: MIT Press.

Hummel, J. E. (2010). Symbolic vs. associative learning. *Cognitive Science*, *34*, 958-965.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99*, 480-517.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220-264.

Keil, F. C. (1986). Conceptual domains and the acquisition of metaphor. *Cognitive Development*, *1*, 73-96.

Keil, F. C. (1989), *Concepts, kinds and cognitive development.* Cambridge, MA: MIT Press.

Kittur, A., Holyoak, K. J., & Hummel, J. E. (2006). Using ideal observers in higher-order human category learning. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twenty Eighty Annual Conference of the Cognitive Science Society* (pp. 435-440). Mahwah, NJ: Erlbaum.

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2004). Feature- vs. relation-defined categories: Probab(alistic)ly Not the Same. In K. Forbus, D. Gentner & T. Regier (Eds.), Proceedings of the Twenty sixth Annual Conference of the Cognitive Science Society (pp. 696-701). Mahwah, NJ: Erlbaum.

Kittur, A., Hummel, J. E., & Holyoak, K. J. (2009). Learning and Inference with Relational Versus Featural Categories. Unpublished manuscript. Carnegie Mellon University.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 1083-1119.

Levene, H. (1960). *Robust Tests for Equality of Variances*, in Contributions to Probability and Statistics, ed. I. Olkin, Palo Alto, CA: Stanford Univ. Press.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130,* 3-28.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50,* 315-353.

Malt, B. C. (1991). Word meaning and word use. In P. Schwanenflugel (Ed.), *The psychology of word meanings.* Hillsdale, NJ: Erlbaum.

Markman, A.B., & Stilwell, C.H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence, 13*, 329-358.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 179-195). New York: Cambridge, University Press.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language And Perception*. Cambridge, MA: Harvard University Press.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Paivio, A. (1971). Imagery and Verbal Processes. New York: Holt, Rinehart and Winston.

Rehder, B. & Ross, B.H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1261-1275.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. O. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York: Cambridge University Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Ross, B. H., & Spalding, T. L. (1994). Concepts and categories. In R. J. Sternberg (Ed.), *Thinking and problem solving* (pp. 119-148). San Diego, CA: Academic Press, Inc.

Shiffrin, R. M., & Styvers, M. (1997). A model for recognition memory: REM--retrieving effectively from memory. *Psychonomic Bulletin & Review*, 145-166.

Smith, E. E., & Medin, D. M. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Wisniewski, E. J. & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology, 39*, 208-238.

Wittgenstein, L. (1953). *Philosophical investigations.* Oxford, UK: Blackwell.