COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2014) 1–33
Copyright © 2014 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12199

Making Probabilistic Relational Categories Learnable

Wookyoung Jung, John E. Hummel

Department of Psychology, University of Illinois

Received 30 October 2013; received in revised form 9 May 2014; accepted 16 June 2014

Abstract

Theories of relational concept acquisition (e.g., schema induction) based on structured intersection discovery predict that relational concepts with a probabilistic (i.e., family resemblance) structure ought to be extremely difficult to learn. We report four experiments testing this prediction by investigating conditions hypothesized to facilitate the learning of such categories. Experiment 1 showed that changing the task from a category-learning task to choosing the “winning” object in each stimulus greatly facilitated participants’ ability to learn probabilistic relational categories. Experiments 2 and 3 further investigated the mechanisms underlying this “who’s winning” effect. Experiment 4 replicated and generalized the “who’s winning” effect with more natural stimuli. 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.

Keywords: Relational category learning; Family resemblance; Higher order relations; Relational invariants; Who’s winning task

1. Introduction

Relational concepts are concepts that specify the relations between things, rather than just the literal features of those things. Such concepts play a central role in human cognition (Gentner, 1983; Holyoak & Thagard, 1995) and lie at the heart of our capacity for language, mathematics, science, art, and almost everything else uniquely human (Penn, Holyoak, & Povinelli, 2008). The acquisition of relational concepts also constitutes a major component of human cognitive development (see Doumas, Hummel, & Sandhofer, 2008, for a review). By adulthood, the average person has mastered hundreds of relational concepts (Asmuth & Gentner, 2005; Goldwater & Markman, 2011; Goldwater, Markman,

Correspondence should be sent to Wookyoung Jung, Department of Psychology, 603 E. Daniel St., Champaign, IL 61820. E-mail: jung43@illinois.edu

& Stilwell, 2011; Markman & Stilwell, 2001): We all know that a *barrier* is something that stands between one thing and another; that a *conduit* is something that transports something else (water, electricity, karma) from one place to another; that a *friend* is someone who likes and is liked by another. Even a concept as simple as *mother* is defined, not by any specific features, but by the relation between the mother and her child (e.g., what features does a human mother share with, say, a mother tarantula? see Markman & Stilwell, 2001). As common and seemingly simple as these concepts are, they are in fact all quite abstract: A barrier can be a piece of concrete, a chasm, or even a person's race or attitudes.

Given the importance of relational concepts in human mental life, an important question is how such concepts are acquired: How do we come to know what a barrier is, or what *larger-than* means? This question is complicated by the fact that, at least by adulthood, many of our relational concepts are largely independent of (i.e., invariant with) their arguments (Doumas et al., 2008; Hummel & Holyoak, 1997, 2003): We understand that *larger-than* means the same thing in the statement "Jupiter is larger than Saturn" as in the statement "The nucleus of an atom is larger than the electrons," even though Jupiter and Saturn are very different than atomic nuclei and electrons. And we know that a *barrier* is, in some abstract sense, the same whether it is a concrete structure that impedes the flow of traffic on a road, or a person's inability to afford tuition at an elite university.

This kind of argument-invariance poses a difficulty for acquiring relational concepts because, although we eventually come to understand relations as distinct from their arguments, we never actually get to experience relations disembodied from their arguments: No one has ever seen an instance of *larger-than* without some specific thing that was larger than some specific other thing. The argument-invariance of relational concepts poses a problem for learning because it implies that associative learning is formally too weak to explain the acquisition of relational concepts (Chomsky, 1959; see also Doumas et al., 2008; Hummel, 2010; Hummel & Holyoak, 1997, 2003), a fact that may help to explain why, of all the primate species, humans appear to be the only one to acquire them (Penn et al., 2008). Transitive inference-like behavior has been observed in non-human animals and has been cited as an instance of relational thinking in these animals (see, e.g., Lazareva, 2012). For example, trained to choose stimulus A over B and B over C, many animals will spontaneously choose A over C. Inasmuch as this kind of behavior depends on explicitly relational thinking, it would appear to be a counterexample to the claim that associative mechanisms are inadequate for relational learning. However, the examples of transitive inference that have been observed in non-human animals can be understood—and have been modeled (see, e.g., McGonigle & Chalmers, 2001)—as instances of straightforward associative generalization (see also Hummel & Holyoak, 2001).

In response to the inadequacy of associative learning to explain the acquisition of relational concepts, some researchers have proposed that relational concepts, including both full-blown schemas (e.g., Gick & Holyoak, 1983; Hummel & Holyoak, 2003) and individual relations, such as *larger-than* (e.g., Doumas et al., 2008), are learned by a process of *structured intersection discovery*. The basic idea is that two situations are structurally aligned, making the correspondences between their parts explicit (Doumas et al., 2008; Gentner, 1983; Gick & Holyoak, 1980, 1983; Hummel & Holyoak, 2003). For example, in the case of the *barrier* concept, one may observe that some person has a goal (to get from point A to point B along a road, or to attend an elite university), which is thwarted by some external reality (a concrete structure between A and B, or an inability to afford tuition at the university), and notice the analogy between them (mapping the traveler onto the hopeful university student, point B along to the road to the elite university, and the concrete structure onto the hopeful student's economic reality), and inducing a schema by retaining the intersection of the two examples: In this case, that some person has some goal that is thwarted by some external reality.

Structural alignment is more powerful than associative learning because it reveals—and depends on—relational matches rather than simple featural matches.[1] In the previous example, the hopeful student's financial reality shares no literal features with the concrete structure blocking the traveler's transit from A to B. Instead, they correspond to one another only by virtue of their shared roles with respect to the actors' goals. Because it relies on the machinery of structural alignment, learning by intersection discovery is equipped to support inferences that extend beyond the statistical regularities of the features of involved in the examples (see Doumas et al., 2008; Gentner, 1983; Hummel & Holyoak, 2003).

In contrast to relational concepts, which are too complex to learn as simple associations, featural concepts (i.e., concepts defined by their exemplars' features, rather than by relations) can be learned associatively. For example, if members of category X tend to have features A1, B1, and C1, whereas members of category Y tend to have features A2, B2, and C2, then it is possible to discriminate Xs from Ys simply by learning associative links (e.g., weighted connections in a connectionist network, or associative links as learned by the Rescorla–Wagner [1973] model) from A1, B1, and C1 to X and from A2, B2, and C2 to Y. That is, there is a good reason to believe that relational and featural concepts require very different learning algorithms: Intersection discovery (or some other algorithm that exploits the machinery of structure mapping) in the case of relational concepts versus simple association in the case of featural concepts (see also Doumas et al., 2008; Hummel, 2010; Hummel & Holyoak, 2003).

To the extent that different learning algorithms underlie the learning of relational and featural concepts, then conclusions drawn from experiments on one kind of category learning may not necessarily apply to the other kind. One of the most robust and replicable conclusions from the literature on category learning from the 1970s to the present (e.g., Kruschke, 1992; Kruschke & Johansen, 1999; Markman & Maddox, 2003; Minda & Smith, 2011; Rosch & Mervis, 1975; Shiffrin & Styvers, 1997; Smith & Medin, 1981) is that people easily learn categories with a *family resemblance*, that is, *probabilistic* structure. In a category with a family resemblance structure, there is no single feature shared by all members of the category. Rather, features tend to occur probabilistically, and "good" members of the category (i.e., members closer to the prototype) tend to have more features in common with other members of the category than "bad" members. The observation that people easily learn categories with a family resemblance structure leads naturally to the conclusion that our natural concepts also have a family resemblance structure, as famously suggested by Wittgenstein (1953).

However, as observed by Kittur and colleagues (Kittur, Holyoak, & Hummel, 2006; Kittur, Hummel, & Holyoak, 2004), one limitation of this conclusion is that all the experiments demonstrating our ability to learn probabilistic category structures were performed using feature-based categories. If feature-based categories are learned associatively, then they should be easily learnable even if they have a probabilistic structure (provided the features are sufficiently predictive of category membership). But if relational categories resist learning by association[2]—and in particular, if they are learned by a process akin to structured intersection discovery—then they should not be learnable when they have a family resemblance structure: If there is no relation that all members of a relational category have in common, then the intersection of the category's exemplars will be the empty set. That is, the intersection-discovery theory of relational learning predicts that probabilistic relational categories ought to be (virtually) unlearnable. In several experiments, Kittur and colleagues demonstrated that, indeed, they appear to be.

In their first experiment, Kittur et al. (2004) used a two-by-two design, crossing featurally versus relationally defined categories with probabilistic versus deterministic structures. They found that, although participants had no difficulty learning deterministic structures, whether they were featural or relational, and had no difficulty learning featural categories, whether they were deterministic or probabilistic, they had great difficulty learning probabilistic relational categories. Indeed, roughly half of their participants never reached criterion even after 600 trials of exposure to the category members. Kittur and colleagues concluded that participants in the relational conditions used a form of intersection discovery to learn the categories, and that this strategy failed catastrophically when the categories had a probabilistic structure.

Although Kittur et al.'s (2004) conclusion is consistent with their findings, it remains possible that relational categories are simply harder to learn than featural ones and that probabilistic categories are harder to learn than deterministic ones and that, put together, these two sources of difficulty interact to make probabilistic relational categories especially difficult. That is, perhaps people learn featural and relational structures in the very same way(s), but relational structures are simply harder to acquire (but see Tomlinson & Love, 2010; Goldwater & Markman, 2011; Goldwater et al., 2011).

However, the prior literature on relational category learning suggests that this simple *conjunction of two difficulties* account may not be correct. In particular, consistent with the intersection discovery account, this literature suggests that relational category learning might be characterized by an attempt to discover an invariant that unifies members of a relational category. Murphy and Allopenna (1994) and Rehder and Ross (2001) showed that category structures that map onto learner's existing schemas are easier to acquire than those that do not. It is possible that this mapping supplies an invariant that unifies members of the relational category. Similarly, *theory-based* categories (e.g., Ahn & Luhmann, 2004; Carey, 1985; Gelman, 2003; Heit, 1994; Keil, Smith, Simons, & Levin, 1998; Medin, 1989; Murphy & Medin, 1985; Rehder, 2003)—a variety of relational category—often have the property that their members may lack any obvious first-order similarities but are united by virtue of conforming to a higher order relational (often goal-based) schema. For example, the members of Barsalou's (1983) ad hoc category *items to take out of a house during a fire* share few or no obvious features but are instead united primarily by their joint membership in the category. A man who jumps fully clothed into the pool at a party may be categorized as *drunk* even if he shares no obvious first-order features with other drunk people (Murphy & Medin, 1985). And Rehder and Ross (2001) showed that categories that are rendered coherent by virtue of their members sharing an unstated higher order relation (namely, that their stated first-order properties work together to achieve a goal) are easier to learn than those that are not. Accordingly, we reasoned that, faced with the task of learning probabilistic relational categories, anything that encourages the learner to discover a higher order relation that remains invariant over members of a category—effectively rendering the category deterministic—ought to substantially facilitate learning.

In Experiments 1–3, participants learned categories of simple "objects," each composed of a circle and a square. (Experiment 4 used artificial "cells" rather than circles and squares, but the category structures were isomorphic to those used in Experiments 1–3.) Following Kittur et al., the categories were relational in the sense that category membership was determined by the relative size, darkness, and locations of the circle and square: In the prototype of category A, the circle was *larger*, *darker*, *above*, and *in-front* of the square; in the prototype of B, the circle was *smaller*, *lighter*, *below*, and *behind* the square (see Fig. 1). Each exemplar of A and B shared three relations with its own prototype and one with the prototype of the opposite category (e.g., one exemplar of A had a circle that was *larger*, *darker*, *above,* and *behind* the square). The categories were probabilistic in the sense that there was no single relation that remained invariant over all exemplars of a category. Although these stimuli lack ecological validity, the use of highly abstract stimuli in studies of category learning has a long and distinguished history, going back at least as far as Bruner, Goodnow, and Austin (1956) and extending at least as recently as Tomlinson and Love (2010), spanning every year in between. In addition, Experiment 4 replicated Experiments 1–3 using more natural "cell" stimuli, and Jung and Hummel (2011) conducted very similar experiments with more natural "insect" stimuli.

Although none of the relevant first-order relations remained invariant over all exemplars of a category, there is an unstated higher order relation that does remain invariant over the members of a category. Specifically, in all members of category A, the circle will stand in three out of four of the relations *larger*, *darker*, *above*, and *in-front* with respect to the square; in all members of B, the square will stand in three out of the four relations *larger*, *darker*, *above*, and *in-front* with respect to the circle. By analogy to a game, if being *larger* gets a shape a "point," being *darker* gets a point, being *above* gets a point, and being *in-front* gets a point, then in all members of category A, the circle is "winning" (with three out of the four possible "points") and in all members of B, the square is "winning." Whether the circle or square is "winning" could thus serve as a higher order (in the sense that it is a relation defined over other relations) invariant that remains true of all members of a category, and which serves to distinguish, deterministically, members of one category from members of the other. As such, we hypothesized

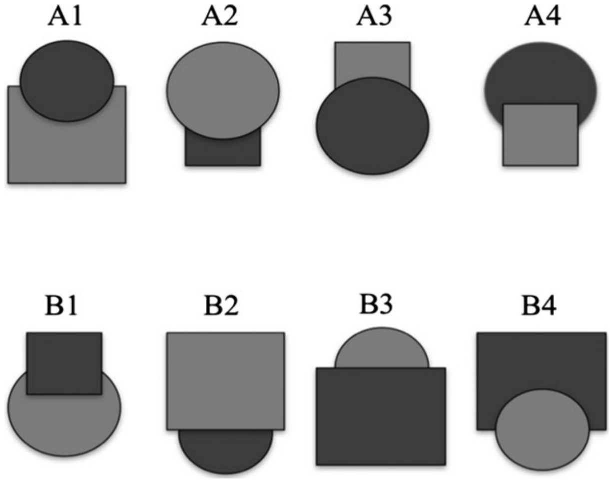

Fig. 1. Exemplars of category A and B used in Experiments 1–3. Exemplars of each category were made by switching the value of one relation in the prototype.

that by changing the task from "learn which exemplars belong to A versus B" to "learn whether the circle or square is winning," we might render otherwise identical category structures learnable.

We report four experiments testing and supporting this hypothesis.

2. Experiment 1

Experiment 1 served as an initial investigation of the schema induction hypothesis, along with two alternative hypotheses about what might make probabilistic relational categories difficult to learn. The first alternative hypothesis is that participants may simply be biased against learning relational categories. As a test of this hypothesis, Experiment 1a manipulated, between subjects, how explicit the instructions were about the fact that the categories were relational. One group of participants was explicitly informed that the categories were defined in terms of whether the circle or square was *larger*, *darker*, *above*, or *in front* of the square, or vice versa[3]; in the other condition, the relations were not explicitly named in the instructions and participants were not informed that the relations between the circle and square (as opposed, say, to their literal features) were relevant to category membership.

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 participants to adopt a suboptimal learning strategy. To test this hypothesis, Experiment 1a also varied whether participants were informed about the probabilistic category structure. Participants in the *clue* condition were explicitly informed that no single property would always work as the basis for categorizing the exemplars. Participants in the *no clue* condition were given no such clue. To the extent that participants 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.

Finally, to test the hypothesis that relational categories are learned by a process akin to schema induction, Experiment 1a manipulated participants' nominal learning task. Participants in the category-learning condition were instructed to learn whether each exemplar belongs to category A or B; those in the "who's winning" condition were instructed to press the A key "if the circle is winning" or to press the B key "if the square is winning." Aside from these differences in the instructions, the stimulus-response mappings were identical across all conditions. That is, any stimulus in which the circle is "winning" also belongs to category A and any stimulus in which the square is winning belongs to B. Although it may seem odd or ecologically invalid to ask participants to learn whether a circle or a square is "winning," that this task *is* learnable is revealed by our data; and what is more important than its ecological validity is the question of whether it is *more* learnable than "does this stimulus belong to A or B?": Natural or not, the intersection discovery hypothesis predicts that it will be. These three variables—relations named versus not, clue provided versus not, and category learning versus "who's winning" task—were manipulated orthogonally, resulting in a total of eight experimental conditions.

Finally, one potential explanation of the difficulty participants had learning Kittur et al.'s probabilistic relational categories concerns the nature of the relations themselves. Many or most relational concepts are defined by *between-object* relations—that is, the relations between the categorized object and some other object(s): A mother is the mother of a child; a barrier blocks the path between a second object and a third; and a conduit carries a second object from a third object (or location) to a fourth. By contrast, the categories used by Kittur and colleagues, and in most of the experiments reported here, are defined by *within-object* relations—relations between the parts of a single category member. Perhaps it is simply unnatural for participants to learn (especially probabilistic) categories based on within-object relations (but see, e.g., Biederman, 1987; Hummel, 2001; Hummel & Biederman, 1992; Marr, 1982; Saiki & Hummel, 1996, on the importance of within-object relations in object categorization). Experiment 1b tested whether probabilistic relational categories are more learnable when the relevant relations are defined between rather than within the to-be categorized objects.

2.1. Experiment 1a

2.1.1. Method

2.1.1.1. Participants: A total of 153 participants participated in the study for course credit. Each participant was randomly assigned to one of the eight conditions.

2.1.1.2. 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.

2.1.1.3. Design: The experiment used a 2 (*relations named* vs. *not named*) × 2 (*clue* vs. *no clue*) × 2 (*categorize* vs. *who's winning* task) between-subjects design.

2.1.1.4. 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, participants 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 14 of 16 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.

2.1.2. Results

2.1.2.1. 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 the criterion were treated as though they had reached criterion on the last block. Fig. 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—$F(1, 145) = 25.826$, $MSE = 2{,}267{,}729$, $p < .001$, 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*,—$F(1, 145) = 5.98$, MSE = 525,066, $p < .05$, indicating that the effect of providing the clue was more pronounced for relation *not named* (M = 291, SD = 243) than for *relations named* (M = 392, SD = 297). Finally, there was a reliable three-way interaction between *relations named*, *clue*, and *task*—$F(1, 145) = 4.10$, MSE = 359,946, $p < .05$. As shown in Fig. 2, relation naming interacted with the clue differently across the two tasks. With the *who's winning* task, the effect of the clue (M = 180, SD = 219) was roughly equivalent to the effect of naming the relations (M = 182, SD = 250), with each reducing trials to criterion relative to providing no clue (M = 244, SD = 301) or relations *not named* (M = 242, SD = 271). By contrast, for participants given the *categorize* task, naming the relations without providing the clue (M = 315, SD = 258) and providing the clue without naming the relations (M = 382, SD = 261) were both beneficial relative to doing neither (M = 503, SD = 376), although these trends did not reach statistical reliability in our sample, $t(41) = 1.912$, $p = .063$ for *relations named* without the *clue,* $t(38) = 1.157$, $p = .254$ for *relations not named* with the *clue*). However, both naming the relations and providing the clue together did not facilitate category learning (M = 623, SD = 381), and in fact impaired learning relative to either naming the relations—$t(37) = -2.999$, $p < .01$, or providing the clue in isolation, $t(34) = -2.214$, $p < .05$.

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.

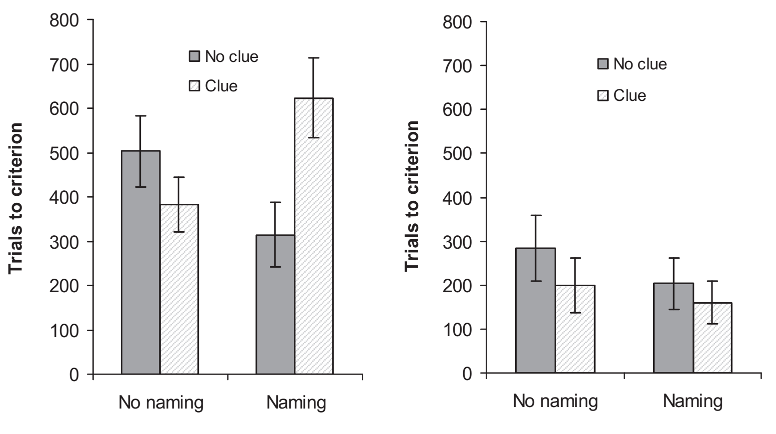

Fig. 2. Mean trials to criterion in the categorize (left) and who's winning (right) conditions in Experiment 1a. Error bars represent standard errors.

2.1.2.2. Response times: We analyzed response times on individual trials in order to gain insight about the strategies participants in *categorize* and *who's winning* may have adopted. There was a reliable main effect of task—$F(1, 145) = 11.280$, $MSE = 12.548$, $p < .01$. Response times in the *categorize* condition ($M = 2.17$, $SD = 1.12$) were reliably shorter than in the *who's winning* condition ($M = 2.75$, $SD = 1.06$). The interaction between *task* and *clue* was reliable—$F(1, 145) = 3.955$, $MSE = 4.400$, $p < .05$, such that for the categorize condition, providing the clue ($M = 2.00$, $SD = 0.94$) made response times shorter than providing no clue ($M = 2.32$, $SD = 1.26$)—$t(77) = 1.245$, $p = .217$, whereas for the who's winning condition, providing the *clue* ($M = 2.93$, $SD = 1.09$) made response times longer than providing no *clue* ($M = 2.57$, $SD = 1.02$)—$t(72) = -1.465$, $p = .147$. Finally, there was a reliable three-way interaction between *relations named*, *clue*, and *task*—$F(1, 145) = 9.522$, $MSE = 10.593$, $p < .01$. As shown in Fig. 3, relation naming interacted with the clue differently across the two tasks. With the *categorize* task, the pattern of response times is the complement of the pattern of accuracy: Response times were numerically slowest when the relations were named and the clue was not given ($M = 2.67$, $SD = 1.49$) and numerically fastest when the relations were named and the clue was given ($M = 1.65$, $SD = 0.90$); response times fell in between, albeit with the reversed pattern, when the relations were not named. In contrast, for participants given the *who's winning* task, naming the relations and providing the clue produced the longest response times ($M = 3.31$, $SD = 1.27$). Without naming the relations, response times were equivalent regardless of providing the clue ($M = 2.55$, $SD = 0.72$ with the clue, $M = 2.54$, $SD = 1.04$ without the clue). Thus, with the *who's winning* task, as with the *categorize* task, the pattern of trial-by-trial response times is roughly the complement of the pattern of trials to criterion.

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

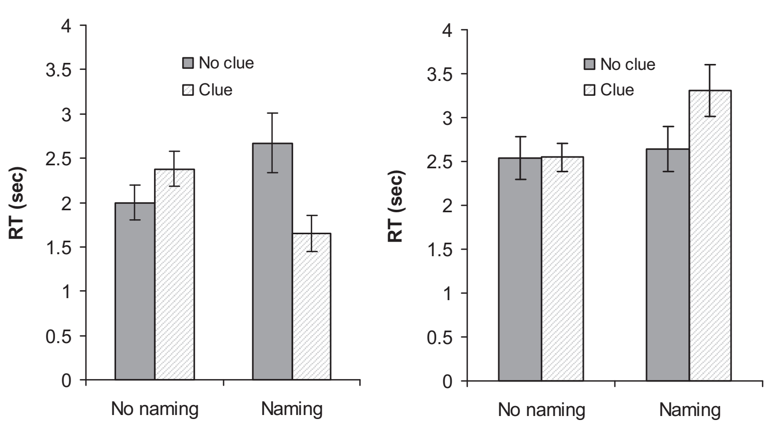

Fig. 3. Mean response times in the categorize (left) and who's winning (right) conditions in Experiment 1a. Error bars represent standard errors.

the latter were attending to the exemplars' relations, including, potentially, higher order relations.

2.2. Experiment 1b

Like the experiments of Kittur and colleagues, Experiment 1a required participants to learn categories defined by the relations between the parts of a single object (i.e., *within-object* relations). That is, the circle and square were described to participants as parts of a single object, whose spatial relations were relevant to category membership. By contrast, many relational categories, including role-governed categories (Markman & Stilwell, 2001), are defined by *between-object* relations—relations between the to-be-categorized object and some other object(s) (e.g., *mother*, *barrier*, *friend*, etc.). Experiment 1b served as a test of whether this within- versus between-object distinction might account for the difficulty participants had learning the relational categories in the categorize conditions of Experiment 1a. Specifically, Experiment 1b explored whether probabilistic relational categories are easier to learn when the relevant relations are defined between- rather than within-objects. In the within-object condition of this experiment, participants, like those in Experiment 1a, were told they were learning to categorize objects composed of two parts (a circle and a square), where the spatial relations between the parts were relevant to category membership. In the between-object condition of this experiment, participants were instructed to categorize either the circle or the square (counterbalanced) in terms of its relations to the other object (i.e., square or circle, respectively). That is, participants in the between-object condition were encouraged to regard the circle and square as separate objects, and thus the relevant relations as between-object relations.

2.2.1. Method

2.2.1.1. Participants: Forty eight undergraduates participated in the experiment for course credit. Each participant was randomly assigned to one of the three conditions.

2.2.1.2. Materials: The stimuli were the same as those used in Experiment 1a.

2.2.1.3. Procedure: Participants were randomly assigned to one of three conditions: *Within-object*, in which their task was to categorize the circle-and-square stimulus as a single object; *between-objects*, in which their task was to categorize either the circle or the square (counterbalanced) based on its relations to the other object; and who's winning. The procedure was otherwise identical to the *relations named* and *clue given* condition of Experiment 1a.

2.2.2. Results

2.2.2.1. Trials to criterion: Fig. 4 shows trials to criterion as a function of condition. A 3 (*within-object* vs. *between-object* vs. *who's winning*) between-subjects design ANOVA revealed a main effect of task—$F(2, 45) = 8.867$, $MSE = 80{,}740.622$, $p < .01$. Participants

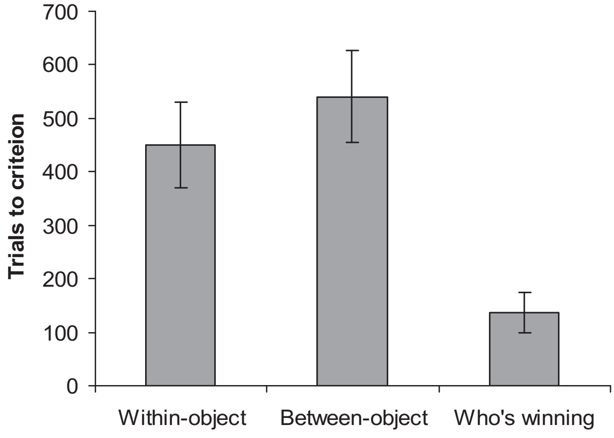

Fig. 4. Mean trials to criterion by condition in Experiment 1b. Error bars represent standard errors.

in *who's winning* (M = 137, SD = 147) reached criterion in fewer trials than in *within-object* (M = 450, SD = 319; by Tukey's HSD, $p < .01$) and *between-object* (M = 540, SD = 345; by Tukey's HSD, $p < .01$). There was no reliable difference between performances in the *within-* and *between-object* conditions (by Tukey's HSD, $p < .646$).

2.2.2.2. Response times: As in Experiment 1a, we analyzed response times on individual trials (Fig. 5). There was a reliable effect of task—$F(2, 44) = 14.524$, $MSE = 22.712$, $p < .001$, such that RTs in *who's winning* (M = 3.55, SD = 2.04) were reliably longer than in both *within-object* (M = 1.55, SD = 0.42; by Tukey's HSD, $p < .001$) and *between-object* (M = 1.43, SD = 0.60; by Tukey's HSD, $p < .001$). There was no reliable different between *within-object* and *between-object* (by Tukey's HSD,

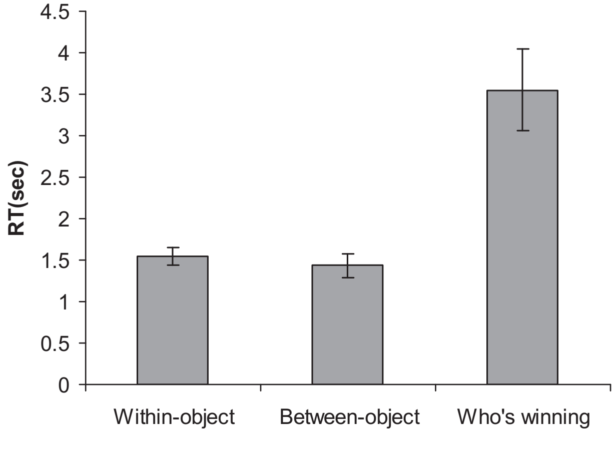

Fig. 5. Mean response times by condition in Experiment 1b. Error bars represent standard errors.

p = .964). Experiment 1b thus showed a speed-accuracy trade-off similar to that observed in Experiment 1a.

2.2.3. 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. Experiment 1a showed that, for participants 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 of Experiment 1a 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 (1.65 s per trial) and slowest in the *who's winning, relations named*, and *clue* condition (3.31 s per trial). Inasmuch as features can be perceived and encoded faster than relations, response times on individual trials are expected to be faster for participants who are responding to features than for those who are responding to relations. 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 (e.g., "dark," "large," "grey," etc.) rather than dimensions ("darkness," "size," "placement") or relations ("darker," "larger," "in front") 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 1a's *categorize, relations named*, and *clue* condition may have abandoned the use of the stated first-order relations (*larger*, *darker*, *above*, and *in front*) 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. In other words, upon being told that a deterministic relational learning strategy wound not work, rather than abandoning their assumption that relational categories should be deterministic, these participants may instead have abandoned the idea that they were relational.

Experiment 1b investigated whether the results of Experiment 1a, and those of Kittur and colleagues, can be attributed to the fact that our stimuli, like those of Kittur et al., defined the object categories in terms of within- rather than between-object relations and showed that this variable had little or no effect on participants' ability to learn the categories (in fact, learning was numerically worse in the between- than within-object condition of this experiment). Together, the results of Experiment 1 suggest that relational categories are learnable only if category members possess one or more relations that remain invariant across members of a category or if, like the "who's winning" condition of Experiment 1, the category-learning task is structured in a way that makes it possible to discover an invariant higher-order relation.

Experiment 2 further investigated the nature of this "who's winning" effect by comparing learning in the "who's winning" task to learning in related tasks. Our second study also provided an opportunity to replicate the basic findings of Experiment 1.

3. Experiment 2

The most striking result of Experiment 1 was the main effect of *who's winning* versus *categorize*. Accordingly, Experiment 2 sought to further elucidate the reasons for this effect. 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 participants to compare the circle and square in some manner that the category-learning task does not. For example, perhaps participants 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. Although this interpretation is challenged by the results of Experiment 1b, it remains possible that explicitly labeling the circle–square relation (i.e., as a winning/losing) in "who's winning" facilitated discovery of that relation in a way that simply categorizing the figures as separate objects in the between-object task of Experiment 1b did not. On this account, any task that encourages participants to learn a relation between the circle and square ought to facilitate learning, whether the to-be-learned relation has any prior meaning or not. For example, asking participants "who's daxier?" should encourage the same kind of comparison as "who's winning." If it is the comparison process that is responsible for the "who's winning" effect, then the "who's daxier" task should result in a comparable improvement (relative to categorization) to "who's winning." But to the extent that the "who's winning" task facilitates performance in part by invoking reasoning schemas for winning/competition, then "who's daxier" may facilitate performance relative to the categorize task, but it ought not facilitate it as much as "who's winning" does.

Our second hypothesis was that a schema for what "winning" consists of may facilitate learning by encouraging participants 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 require participants to discover relations between the circle and square.

Where these hypotheses make divergent predictions is in the potential role of *consistent* versus *mixed role assignment* in the effect. The instructions refer to the relevant relations by naming one role of each relation: Participants 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*. In Experiment 1, the assignment of relational roles to categories was *consistent*, in the sense that all the roles named in the instructions were assigned to the circle in category A (with the unnamed roles assigned to the square) and all the roles not named in the instructions were assigned to the circle in category B (with the named roles assigned to the square). 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 facilitatory effect of the winning task is due to the involvement of a "winning" schema, per se, then having all the named roles consistently assigned to a single shape within a category (as in Experiment 1) ought to lead to faster learning than having half the winning roles assigned to one shape and half to the other in each category. By contrast, to the extent that the effect of "who's winning" simply reflects the role of comparison, then consistent versus mixed role assignment 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?", an additional effect of consistent versus mixed role assignment.

Experiment 2 tested both hypotheses by orthogonally crossing task (*categorize* vs. *who's daxier* vs. *who's winning*) with role assignment (*consistent* vs. *mixed*). 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.

3.1. Method

3.1.1. Participants

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

3.1.2. Materials

There were two types of stimuli: In the *consistent role assignment* condition, the prototypes were identical to those of Experiment 1. In the *mixed role assignment* condition, the named roles were mixed across the two shapes within the members of A and B (*categorize* condition), the "daxier" shape (*daxier* condition) or the "winning" shape (*winning* condition). The precise mixing of roles was counterbalanced: In one case, the circle in 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).

3.1.3. Design

The experiment used a 3 (*categorize* vs. *who’s daxier* vs. *who’s winning task*) × 2 (*consistent* vs. *mixed role assignment*) between-subjects design.

3.1.4. 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.

3.2. Results

3.2.1. 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 Fig. 6. A 3 (*categorize* vs. *daxier* vs. *winning*) × 2 (*consistent* vs. *mixed*) between-subjects design ANOVA revealed a main effect of task—$F(2, 99) = 11.352$, $MSE = 1{,}158{,}433$, $p < .001$. 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, $p < .01$). Participants given the *who’s daxier* task ($M = 492$, $SD = 318$) took reliably fewer trials to reach criterion

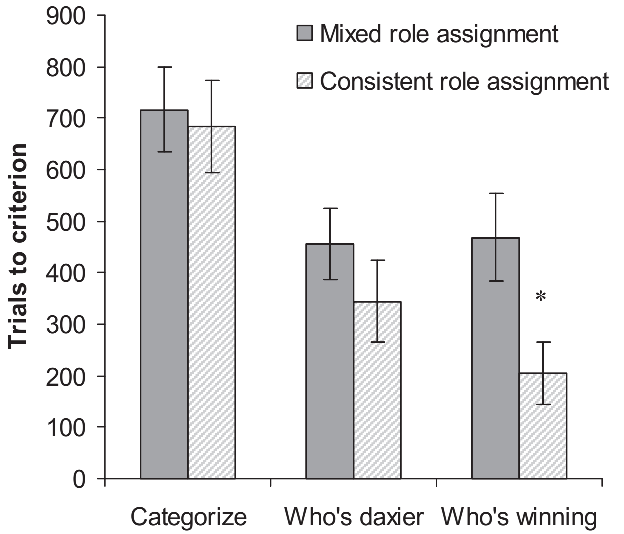

Fig. 6. Mean trials to criterion by condition in Experiment 2. Error bars represent standard errors. *$p < .05$.

than those in the *categorize task* (M = 699, SD = 317; by Tukey's HSD, $p < .05$). Participants given the *who's winning* task took fewer trials to reach criterion as those given *who's daxier* (by Tukey's HSD, $p < .05$). There was also a reliable main effect of role assignment—$F(1, 99) = 4.701$, MSE = 479,678, $p < .05$. As expected, participants in the consistent conditions (M = 381, SD = 360) reached criterion faster than those in the mixed conditions (M = 521, SD = 341). This difference between consistent (M = 206, SD = 279) and mixed (M = 468, SD = 359) was reliable only in the *who's winning* condition—$t(36) = -2.534$, $p < .05$.

3.2.2. Response times

As in Experiment 1, we analyzed response times on individual trials (Fig. 7). There was a reliable effect of task—$F(2, 99) = 9.296$, MSE = 13.588, $p < .001$, 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, $p < .001$) and RTs in *who's daxier* (M = 2.55, SD = 1.18) were reliably longer than in *categorize* (by Tukey's HSD, $p < .05$). The main effect of *consistent* (M = 2.56, SD = 1.45) versus *mixed* (M = 2.36, SD = 1.14) role assignment was not reliable—$F(1, 99) = 0.402$, MSE = 0.588, p = .527. Experiment 2 thus showed a speed-accuracy trade-off similar to that observed in Experiment 1.

3.3. 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 consistent and mixed conditions is consistent with

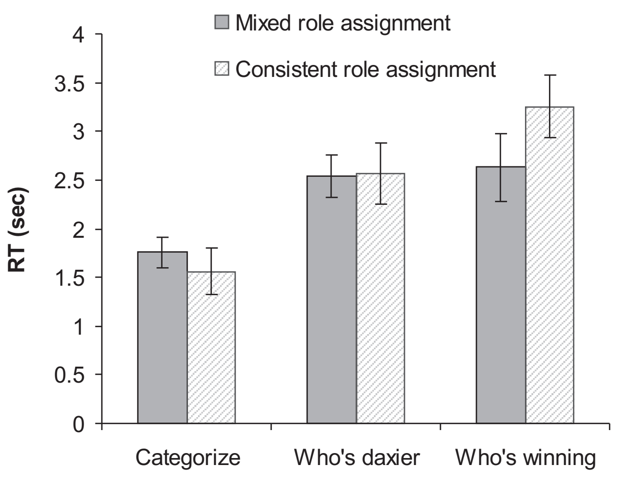

Fig. 7. Mean response times in the categorize, who's daxier, and who's winning conditions in Experiment 2. Error bars represent standard errors.

the hypothesis that *who's winning* (like *who's daxier*) encourages participants to compare the circle and square in a way that categorization does not. This hypothesis is further supported by the fact that participants in the *mixed 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 participants in the *consistent winning* condition learned faster than those in either the *mixed 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.

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 facilitates finding such an invariant with our stimuli. Experiment 2 also demonstrated that simply having a task that encourages participants to think of the circle and square as separate objects (the *who's daxier* task) is not, by itself, sufficient to achieve the same facilitation enjoyed by participants 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 participants' superior performance in the *who's winning* condition.

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. 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 participants 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. Participants 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 participants 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 participants to discover a higher order relation between the circle and the square. And like *who's winning*, but unlike *who's daxier*, the roles of this relation have opposite valence (presumably it is "good" to be liked by Britney) and it has meaning. A fifth group of participants 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

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

in valence across its roles (i.e., it is presumably neither good nor bad to be from Nebraska). These properties of our 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 consistent versus mixed role assignment as an additional check on their similarity to *who's winning* or *who's daxier*. To the extent that *Britney* versus *Nebraska* are like *who's winning* versus *who's daxier*, respectively, they should show the same patterns of sensitivity versus insensitivity to role assignment.

4.1. Method

4.1.1. Participants

One hundred and ninety-one undergraduates participated in the experiment to fulfill a course requirement. Each participant was randomly assigned to one of the 10 conditions.

4.1.2. 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.

4.1.3. Design

The experiment used a 5 (*categorize* vs. *who's daxier* vs. *which one comes from Nebraska vs. which one would Britney Spears like* vs. *who's winning*) × 2 (*consistent* vs. *mixed role assignment*) between-subjects design.

4.1.4. Procedure

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

4.2 Results

4.2.1. Trials to criterion

We conducted a 5 tasks (*categorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) × 2 assignments (*consistent* vs. *mixed*) ANOVA, but Levene's test (Levene, 1960) for equality of variance revealed a significant difference in variance across five groups in trials to criterion ($p < .001$). The 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 (*categorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) × 2 assignments (*consistent* vs. *mixed*) between-subjects design ANOVA revealed that Levene's test was not significant ($p = .588$). As shown in Fig. 8, there was a main effect of the task—$F(4, 181) = 9.811$, $MSE = 7.108$, $p < .001$. Since our main interest is in how the task itself affects the learning of probabilistic relational categories, we first report the data from the consistent role assignment condition. A 5 tasks (*categorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) between-subjects ANOVA revealed a reliable effect of the task—$F(4, 106) = 11.150$, $MSE = 7.525$, $p < .001$. As in the previous experiments, participants in *who's winning* ($M = 127$, $SD = 127$) reached criterion faster than those in *who's daxier* ($M = 276$, $SD = 241$; $p < .01$) as well as those in *categorize* ($M = 523$, $SD = 341$; $p < .001$; by Tukey's HSD). Participants in *daxier* took reliably fewer trials to reach criterion than those in *categorize* (by Tukey's HSD, $p < .05$). Performance in *Britney* ($M = 186$, $SD = 155$) was not reliably different than in *winning* (by Tukey's HSD, $p = .381$). Participants in *Britney* reached criterion reliably faster than those in *categorize* (by Tukey's HSD, $p < .001$). The number of trials to reach criterion in *Nebraska* ($M = 250$, $SD = 218$) was reliably less than in *categorize* (by Tukey's HSD, $p < .05$). There was also a main effect of role assignment—$F(1, 181) = 26.769$, $MSE = 19.394$, $p < .001$. The *winning* task—$t(38) = 3.717$, $p < .01$, the *Britney* task—$t(33) = 3.036$, $p < .01$, and the *Nebraska* task—$t(35) = 3.006$, $p < .01$, all of which have semantic content, showed reliable differences between the consistent and mixed conditions. There was

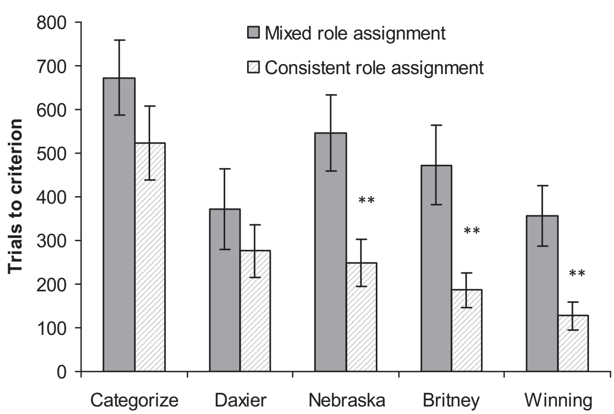

Fig. 8. Mean trials to criterion by condition in Experiment 3. Error bars represent standard errors. **$p < .01$.

no such effect in *categorize*—$t(37) = 1.308$, $p = .199$, or *daxier*—$t(38) = 0.388$, $p = .7$, which lacks semantic content. The interaction between task and role assignment did not reach statistical reliability—$F(4, 181) = 2.208$, $MSE = 1.6$, $p = .07$.

4.2.2. Response times

Levene's test revealed a significant difference in variance across five groups in response times ($p < .05$). The data were log transformed to normalize the skewed distribution. After the transformation, a 5 tasks (*categorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) × 2 role assignments (*consistent* vs. *mixed*) between-subjects design ANOVA revealed that Levene's test was not reliable ($p = .178$). We report response times in the *consistent role assignment* condition since we were mainly interested in how the different tasks affect category learning in that condition. A 5 tasks (*categorize* vs. *daxier* vs. *Nebraska vs. Britney* vs. *winning*) between-subjects design ANOVA revealed a reliable effect of task—$F(4, 106) = 6.177$, $MSE = 0.940$, $p < .001$, (Fig. 9). Participants in *winning* ($M = 3.43$, $SD = 1.73$; by Tukey's HSD, $p < .001$) and *daxier* ($M = 2.89$, $SD = 1.29$; by Tukey's HSD, $p < .05$) 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* ($p = .06$) and RTs in *Nebraska* ($M = 2.69$, $SD = 1.34$) were also marginally longer than RTs in *categorize* ($p = .06$; by Tukey's HSD). As in Experiments 1 and 2, Experiment 3 revealed a speed-accuracy tradeoff.

4.3. 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 participants to consider the circle and square as separate objects. To examine how each of these factors contributes to the acquisition of

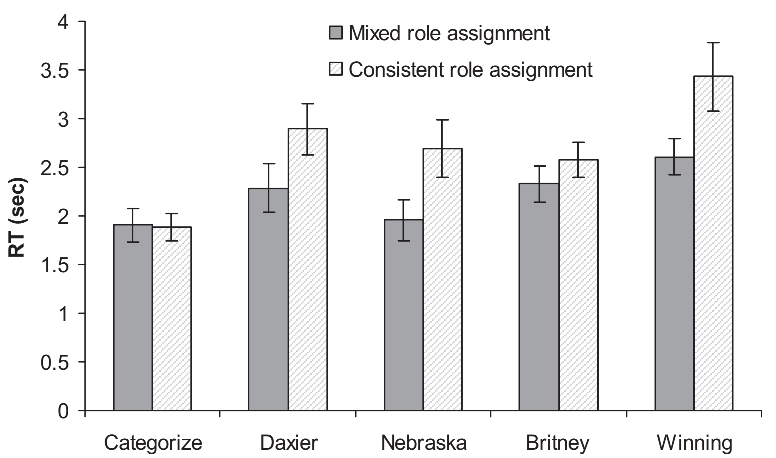

Fig. 9. Mean response times by condition in Experiment 3. Error bars represent standard errors.

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 *daxier* task has no semantic content. Consistent with our assumptions, learning in the resulting conditions fell roughly into three groups: Learning was fastest in *winning* and *Britney*, followed by *daxy* and *Nebraska*, and finally by *categorize*. Within these groupings, differences in rate of learning were not statistically reliable. Between them, differences in learning rate were all reliable, with the exception of the difference between *Britney* and *Nebraska*. This pattern of results suggests that tasks that treat the circle and square as separate objects, have roles with opposite valence, and have semantic content (like *winning* and *Britney*) 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.

5. Experiment 4

The purpose of Experiment 4 was to examine whether the who's winning effect would generalize to stimuli that map more naturally to a real-world task—specifically, a fictional experiment in a biology laboratory. Each stimulus consisted of two "petri dishes" side by side, each containing a number of "cells" (see Fig. 10). The cells in the dishes differed in their number, size, darkness, and elongation. In the prototype of category T (denoted [1,1,1,1]), the cells in the left dish were *more numerous*, *larger*, *darker*, and *more elongated* than those in the right dish (the absolute number, size, darkness, and elongation of the cells in both dishes was free to vary). In the prototype of category V (denoted [0,0,0,0]), the cells in the left dish were *less numerous*, *smaller*, *lighter,* and *rounder* than those in the right dish. As in the previous experiments, each exemplar was constructed by adopting three relations of the corresponding prototype and one of the opposite prototype. For example, in exemplar [0,1,1,1] (category T), the cells in the left dish were *less numerous*, *larger*, *darker*, and *more elongated* than those in the right dish.

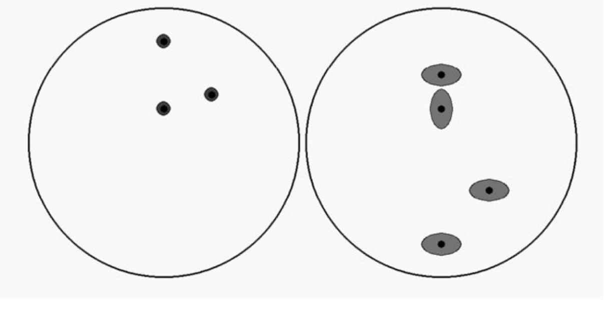

Fig. 10. Exemplar of the cell stimuli used in Experiment 4.

The tasks were constructed to correspond to the critical tasks in Experiments 1–3. In the *categorize* condition of Experiment 4, the participant was asked to categorize the pair of petri dishes as belonging to category T or category V. (The labels T and V, which are on the upper and lower portions of the center of the keyboard, were chosen to avoid stimulus-response compatibility effects with the petri dishes, which were displayed on the left and right sides of the screen.) Participants were informed of the relevant relations between the cells, and they were informed that no single relation would determine category membership by itself. In the *which cells will live longer?* condition, participants were instructed to press the T key if the cells in the left dish would live longer than those in the right, or to press the V key if those in the right dish would live longer. This task was chosen to mimic the *who's winning* task of Experiments 1–3. As in Experiments 1–3, the *categorize* and *live longer* tasks were isomorphic in the sense that any stimulus that would be correctly categorized as a member of T also had the property that the cells on the left would live longer, and any that would be correctly categorized as a member of V had the property that those on the right would live longer (i.e., the stimulus-response mappings were identical across the tasks). We also included a third condition, *which chemical was applied to the cells?*, in which the participants were instructed that either chemical T or chemical V had been applied to the cells in both dishes, and their task was to decide which chemical had been applied. Like the *live longer* task, the stimulus-response mapping in this task was identical to that in the *categorize* task. We included the *which chemical* condition as a more natural (i.e., semantically meaningful) isomorph of the *categorize* condition. Like participants in the *categorize* condition, participants in the *live longer* and *which chemical* conditions were told what the relevant relations were and were informed that no single relation would be perfectly predictive of the correct response.

5.1. Method

5.1.1. Participants

Forty-five undergraduates participated in the experiment for course credit

Each participant was randomly assigned to one of three conditions and none had participated in any of the previous experiments.

5.1.2. Materials

The current experiment used fictional "cells" as stimuli. We designed the cell stimuli to be isomorphic to the circle and square stimuli used in the previous experiments. Each stimulus depicted two "petri dishes" (large circles), each containing between three and eight "cells." The cells in the petri dishes differed in their number (one dish contained more cells than the other), their size (those in one dish were larger in area than those in the other), their darkness (those in one dish were darker than those in the other), and elongation (those in one dish were more elongated than those in the other). In the prototype of category T, the cells in the left petri dish were *more numerous, larger*, *darker*, and *more elongated* than those in the right petri dish. The prototype of category V depicted the opposite relations: The cells in the left petri dish were *less numerous, smaller*, *lighter*, and *rounder* than those in right dish. As in the previous experiments, exemplars of each category were made by switching the value of one relation in the prototype and also two variants of each logical structure were constructed by varying the metric properties size and darkness, providing eight exemplars per category (e.g., category T exemplar [1,0,1,1] would have cells in the left petri dish *more numerous, smaller*, *darker*, and *more elongated* than in the right petri dish).

5.1.3. Procedure

Participants were randomly assigned to one of three conditions: *Categorize*, in which their task was exactly identical to the categorize condition in the previous experiment. Participants were asked to categorize whether a pair of petri dishes belong to category T or V. *Which chemical was applied,* in which their task was to decide which chemical T or V was applied to the petri dishes; or *Which cells will live longer,* in which their task was to decide cells in which petri dish will live longer. All participants in all conditions were informed of the relevant relations and informed that no single relation was perfectly predictive of the correct response.

5.2. Results

5.2.1. Trials to criterion

A 3 (*categorize* vs. *which chemical was applied* vs. *which cells will live longer*) between-subjects design ANOVA revealed a main effect of task—$F(2, 42) = 21.707$, $MSE = 1{,}520{,}230.4$, $p < .001$, (Fig. 11). We performed planned comparisons between all three conditions. Participants in *live longer* ($M = 82$, $SD = 53$) reached criterion in fewer trials than those in *which chemical* ($M = 453$, $SD = 341$)—$t(42) = 3.841$, $p < .001$), and *categorize* ($M = 716$, $SD = 302$)—$t(42) = 6.557$, $p < .001$). Planned comparison between *which chemical* and *categorize* also showed a reliable difference—$t(42) = 2.715$, $p < .05$).

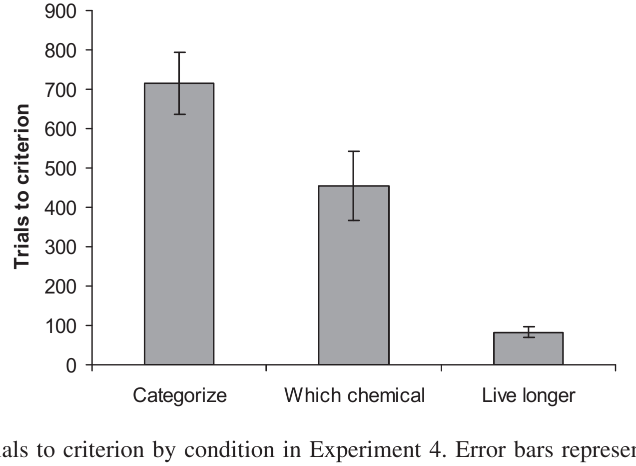

Fig. 11. Mean trials to criterion by condition in Experiment 4. Error bars represent standard errors.

5.2.2. Response times

We analyzed response times on individual trials (Fig. 12). There was a reliable effect of task—$F(2, 42) = 7.191$, $MSE = 3.682$, $p < .01$, such that RTs in *live longer* ($M = 1.96$, $SD = 0.89$) were reliably longer than in *categorize* ($M = 1.04$, $SD = 0.42$; by Tukey's HSD, $p < .01$) and RTs in *which chemical* ($M = 1.81$, $SD = 0.76$) were reliably longer than in *categorize* (by Tukey's HSD, $p < .05$). There was no significant difference between *live longer* and *which chemical* (by Tukey's HSD, $p = .845$). Experiment 4 thus showed a speed-accuracy trade-off similar to that observed in the previous experiments.

5.3. Discussion

Experiment 4 replicated and extended the critical findings of Experiments 1–3. Participants given the *live longer* task learned substantially faster than those given either *categorize* or *which chemical*. If anything the effect of *live longer* was numerically greater (82 trials to criterion) than the effect of *who's winning* (160 trials to criterion in the corresponding condition of Experiment 1) as compared to *categorize* (716 and 623, respectively). The reason for the facilitatory effect of *which chemical* relative to *categorize* is less clear, but it is consistent with the general idea that simply making the task meaningful facilitates participants' learning. In contrast to, say, the *daxy* condition (Experiment 2), which encourages comparison but provides no semantic link between the materials and the task, the *which chemical* task has intuitive semantic relations to differences in cell growth.

6. General discussion

Kittur et al. (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 intersec-

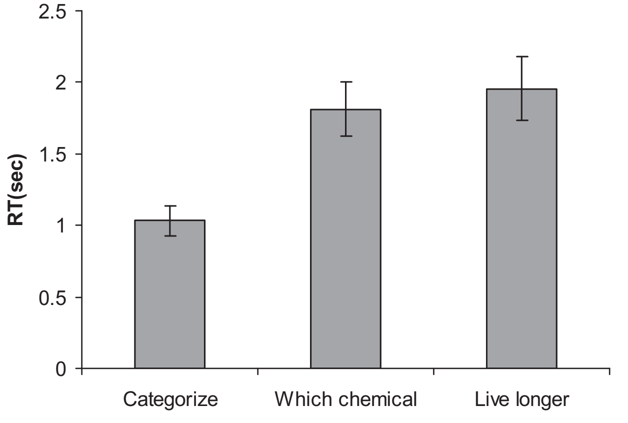

Fig. 12. Mean response times by condition in Experiment 4. Error bars represent standard errors.

tion discovery (Gick & Holyoak, 1983; 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 conditions that render probabilistic relational category structures learnable. Specifically, the intersection discovery hypothesis predicts that any task that leads the learner to discover a (e.g., higher order) property or relation that remains invariant over exemplars of an otherwise probabilistic relational category structure ought to render that structure 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, and they individually facilitated learning in the *categorize* condition, together, they 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 the person that he or she is trying to learn probabilistic relational categories.

Experiment 1 also demonstrated that the difficulty of learning probabilistic relational categories reported here and by Kittur and colleagues does not simply reflect the "unnaturalness" of the stimulus materials used in our and Kittur et al.'s experiments. Although it is arguably unnatural to categorize circle–square pairs based on the relations between them, and this unnaturalness may make the categories difficult to learn, it is at least as unnatural to declare a circle or a square a "winner" based on those same relations, and yet this task is easy to learn.

Experiment 2 demonstrated that this effect is not simply due to participants being encouraged to compare the circle and square as separate objects in the "winning" task: Asking participants "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.

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 the role of a relation whose roles have unequal valence.

Finally, Experiment 4 demonstrated that an isomorphic task, *which cells will live longer?*, using stimuli and responses that map more naturally onto a task the learner may encounter in "real life" (or at least in a biology laboratory), replicates the basic "winning" effect and, if anything, has an even stronger facilitatory effect on learning.

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 those things, 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.

6.1. Relational concepts

Numerous researchers have emphasized the central role of relational structures such as schemas and theories in human cognition (e.g., Barsalou, 1983; Gentner, 1983; Gick & Holyoak, 1983; Murphy, 2002; Murphy & Medin, 1985; Rehder & Ross, 2001). A general consensus that emerges from this literature is that relational concepts behave in a qualitatively different manner than feature-based concepts (e.g., Barsalou, 1983; Doumas et al., 2008; Gentner, 1983; Gick & Holyoak, 1983; Holyoak & Thagard, 1989, 1995; Goldwater & Markman, 2011; Goldwater et al., 2011; Hummel & Holyoak, 1997, 2003; Kittur et al., 2004, 2006; Murphy, 2002; Murphy & Medin, 1985; Rehder & Ross, 2001; Tomlinson & Love, 2010; for a review, see Holyoak, 2005). For example, Goldwater and Markman (2011) showed that analogical comparison increased people's sensitivity to *role-governed categories*—categories such as *friend* or *guest* that are defined by the relational role in which an object is engaged. Goldwater et al. (2011) showed that, when asked to choose words to describe feature-based categories, people tend to choose words describing typical category characteristics; but when asked to choose words describing role-governed categories, people tend to choose words describing ideal characteristics (see also Kittur et al., 2006; Rein, Goldwater, & Markman, 2010). Similarly, Ross and Murphy (1999) observed that people tend to make taxonomic inferences about category membership in response to category members' features, but they are more likely to make script-based inferences about category membership in response to situational (i.e., relational) information. For example, people are more likely to choose a taxonomically related alternative (e.g., another fruit) in response to the question "What food is more likely to contain metacascal?" than to choose a script-based alternative (e.g., another breakfast food). By contrast, when asked to choose which other food is more likely to be used in an initiation ceremony, they are more likely to choose a script-based alternative than a taxonomically related alternative.

The findings reported here suggest that, not only do relational categories behave in a qualitatively different manner than featural categories, they are also *learned* in a qualitatively different manner. In particular, these findings suggest that, as hypothesized here and by Kittur and colleagues, relational categories are learned by a kind of structured intersection discovery, and that this process fails catastrophically with probabilistic category structures. As a result, one of the most robust phenomena observed in the literature on cognitive psychology—prototype effects (see Murphy, 2002; for a review)—do not obtain with relational categories (Kittur et al., 2004).

Our findings also replicate and extend those of Rehder and Ross (2001) with *abstract coherent categories*. Rehder and Ross showed that, when members of a category are united by sharing an unstated higher order relation (specifically, whether the properties of a category member work together in a coherent manner to achieve a goal, such as cleaning spilled oil from water), the categories are easier to learn than when their members do not share this higher order relation. These researchers interpreted their findings in terms of their coherent categories making contact with participants' preexisting schemas in a way that incoherent categories do not. Our results (especially those of Experiments 3 and 4) reinforce the role of preexisting schemas in learners' ability to make sense of and acquire new categories, and extend Rehder and Ross's original findings by highlighting the role of those schemas in the learners' discovery of an invariant that is shared by all members of the relational category.

6.2. The frailty of probabilistic relational concepts

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. 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. 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 "points" more strongly (see also Newell, Dunn, & Kalish, 2010). 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 feature-based concepts, it is inadequate for acquiring relational concepts (see Chomsky, 1959; Deacon, 1997; Doumas et al., 2008; Hummel & Holyoak, 2003) because 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* an infinity of other objects. This observation has led some researchers to hypothesize that relational structures such as schemas (Gick & Holyoak, 1983; 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. 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 it fails catastrophically with concepts that do not have a deterministic structure.

One potential explanation of the data presented here, and by Kittur et al. (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.

7. Conclusion

Four 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.

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. 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 we may have multiple "mother" schemas, as suggested by Lakoff (1987). That is, although "mother" may appear at first blush to be a single (relational) concept without an invariant relational core, it is also possible that the word "mother" refers to multiple concepts, each with its own relational invariants.

The literature on concepts and categories is characterized by a divide between studies emphasizing the family resemblance nature of our concepts and those emphasizing their theory-based, relational nature. One of the most important implications of the current findings (like those of Kittur et al., 2004) is that schemas and theories must contain relational invariants (or else be extremely difficult to acquire). That is, like scientific theories, theory-based concepts may be subject to falsification based on a single compelling counterexample. But also like scientific theories, our theory-based concepts seek to unify observable first-order observations under invariant higher order relations.

Acknowledgments

Portions of this research were presented at the 31st annual conference of the Cognitive Science Society (Amsterdam, the Netherlands). We thank David Landy and Brian Ross for their insights, and Robert Weishappel and Anais Binkowski for their assistance with the experiments. We especially thank Natalie Hummel, then 9 years old, for discovering the “who’s winning” strategy. The research was supported by Air Force Office of Scientific Research Grant FA9550-07-1-0147.

Notes

1. Armed with these relational matches, structure mapping affords complex, relational inferences by a kind of *copy with substitution and generation* (*CWSG*; see Holyoak & Thagard, 1995), in which structures in one situation can be copied into the other situation, substituting the corresponding role bindings (e.g., just as the traveler may try to find an alternate route from A to B around the concrete structure, the hopeful student may try to find an alternate way to pay the university’s tuition). This same kind of CWSG also provides a natural account of abstract, variablized rule following. For example, if one knows the rule *for all x, y, and z, and for all transitive relations, phi, phi(x, y), and phi(y, z) implies phi(x, z)*, and one encounters the example *wickier* (dax, blicket) and *wickier* (blicket, rokum), then one can infer *wickier* (dax, rokum) by binding *wickier* to *phi*, *dax* to *x*, *blicket* to *y,* and rokum to *z*, and applying CWSG. Formally, CWSG is a kind of union discovery, in the sense that it augments the specific example (here, *wickier*, etc.) to become the union of itself and the corresponding, but heretofore missing, elements of the rule. The process of discovering the abstract rule from examples in the first place has been characterized as the complementary process of *intersection* discovery, in which concrete examples are structurally aligned and their relational intersection is taken to be an abstract characterization of both (Doumas et al., 2008; Gentner, 1983; Gick & Holyoak, 1980, 1983; Holyoak & Thagard, 1995; Hummel & Holyoak, 2003). The virtues of this intersection discovery process are that it is based on relational matches, and that it preserves only those abstract structures—presumably, the underlying relational structures—that drive and thus “survive” the alignment and intersection-taking process
2. Although relational concepts such as *larger-than* are formally too complex to learn by association, novel ideas (such as “oranges are larger than grapes”) or categories (such as “jumbo shrimp”) based on familiar relations with novel arguments are learnable by association once the basic relational concepts (here, *larger-than*) are in place (see Doumas et al., 2008; Hummel & Holyoak, 2003). That is, although relations, per se, are not learnable by association, this is not to say that all relational concepts must necessarily defy associative learning. Nonetheless, if the presence of relational concepts invokes a non-associative learning bias on the part of the cognitive architecture, then that bias might be expected to persist even in those cases, such as learning new arguments of familiar relations, when associative learning would otherwise work.
3. Specifically, in this condition, the instructions contained the sentence, “Each object consists of a circle and a square, one of which is larger than the other, one of which is darker, one of which is above the other, and one of which is in front of the other.”

References

Ahn, W., & Luhmann, C. C. (2004). Demystifying theory-based categorization. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 277–300). Mahwah, NJ: Lawrence Erlbaum Associates.

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. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 163–168). Stresa, Italy: Cognitive Science Society.

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

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.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

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.

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, CA: Academic Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. London: Oxford University Press.

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

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

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–28.

Goldwater, M. B., & Markman, A. B. (2011). Categorizing entities by common role. *Psychonomic Bulletin and Review*, *18*, 406–413.

Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, *118*, 359–376.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1264–1282.

Holyoak, K. J. (2005). Analogy. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 117–142). Cambridge, UK: Cambridge University Press.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295–355.

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

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, *8*, 489–517.

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. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Hummel, J. E., & Holyoak, K. J. (2001). A process model of human transitive inference. In M. Gattis (Ed.), *Spatial schemas in abstract thought* (pp. 279–305). Cambridge, MA: MIT Press.

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

Jung, W., & Hummel, J. E. (2011). In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2643–2648). Austin, TX: Cognitive Science Society.

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

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*, 103–135.

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-eighth 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.

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, and Cognition*, *25*, 1083–1119.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lazareva, O. F. (2012). Relational learning in a context of transposition: A review. *Journal of the Experimental Analysis of Behavior*, *97*, 231–248.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics*. Palo Alto, CA: Stanford University Press.

Markman, A. B., & Maddox, W. T. (2003). Classification of exemplars with single- and multiple-feature manifestations: The effects of relevant dimension variation and category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 107–117.

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

Marr, D. (1982). *Vision*. New York: Freeman & Co.

McGonigle, B., & Chalmers, M. (2001). Spatial representation as cause and effect: Circular causality comes to cognition. In M. Gattis (Eds), *Spatial schemas and abstract thought* (pp. 247–277). London: MIT Press.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*, 1469–1481.

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

Minda, J. P., & Smith, J. D. (2011). *Prototype models of categorization: Basic formulation, predictions, and limitations*. (pp. 40–64). New York: Cambridge University Press.

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

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 904–919.

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

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: A state-trace analysis. *Memory and Cognition*, *38*, 563–581.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*, 109–129.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1141–1159.

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

Rein, J. R., Goldwater, M. B., & Markman, A. B. (2010). What is typical about the typicality effect in category-based induction? *Memory & Cognition*, *38*, 377–388.

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

Ross, B. H., & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, *38*, 495–553.

Saiki, J., & Hummel, J. E. (1996). Attribute conjunctions and the part configuration advantage in object category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1002–1019.

Shiffrin, R. M., & Styvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

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

Tomlinson, M. T., & Love, B. (2010). When learning to classify by relations is easier than by features. *Thinking and Reasoning*, *16*, 372–401.

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