

# Similarity and Property Effects in Inductive Reasoning

Evan Heit and Joshua Rubinstein

Three experiments investigated the proposal that inductive inferences about different properties depend on different measures of similarity. In Experiments 1 and 2, Ss were given the premise that a category of animals has some property and judged the probability that another category of animals also has that property. Ss made the strongest inferences when the kind of property (anatomical or behavioral) corresponded to the kind of similarity between the animal categories (anatomical or behavioral). These results cannot be explained in terms of a single measure of similarity underlying induction. In Experiment 3, Ss rated the similarity of animal pairs *with respect to* anatomy or behavior. Regression analyses showed that both behavioral and anatomical similarity influenced behavioral inferences, but only anatomical similarity influenced anatomical inferences.

People use similarity to make inferences. When Category A has Property P, people are more likely to infer that Category B also has Property P to the extent that A and B are perceived to share other features. We propose that, in addition, Property P itself has an important role in determining which features are used to evaluate the similarity between Categories A and B. Inductive reasoning does not rely on a fixed notion of similarity; instead, the similarity between A and B will be evaluated with respect to features that are relevant to Property P. In particular, we propose that different measures of similarity are used when people make inferences about anatomical properties and behavioral properties of animals so that people focus on similarity in terms of anatomical features for anatomical inferences, and they focus on behavioral similarity for behavioral inferences.<sup>1</sup>

For example, the premise that bears have livers with two chambers lends inductive support to the conclusion that whales have livers with two chambers because bears and whales share many internal, anatomical and biological features. But the premise that bears tend to move in circles during cold weather does not lend much inductive support to the conclusion that whales tend to move in circles the same way because bears and whales are not very similar in terms of locomotion and related behaviors. In this example, the property being considered influences which features of bears and whales are important for induction.

Previous studies of inductive reasoning may be divided into two groups. The first group includes studies that described the effects of similarity between premise and conclusion categories on induction, but only for a single property or for a homogeneous set of properties (Osherson, Smith, Wilkie, López, &

Shafir, 1990; Osherson, Stern, Wilkie, Stob, & Smith, 1991; Rips, 1975). Yet the role of properties may be just as important as the role of categories. The second group includes studies that have shown some divergences in inductive reasoning that are based on different properties, especially that people are more willing to make generalizations about some properties than other properties (Carey, 1985; Gelman, 1988; Gelman & Markman, 1986; Nisbett, Krantz, Jepson, & Kunda, 1983; Springer, 1992). Our research is intended to add to these two bodies of evidence by making explicit the distinction between anatomical and behavioral properties and by specifying the roles of two kinds of similarity, anatomical and behavioral, in inductive reasoning.

## Research on Similarity Effects

Some research has contributed to the study of induction by describing structural relations between similarity and induction in detailed mathematical form. What is crucial about these studies is the assumption that inductive reasoning can be accounted for in terms of a single measure of similarity. Although these studies were successful at modeling induction by using just one kind of similarity, they did not attempt to describe reasoning about more than one kind of property. Therefore, these studies simply cannot address the issue of how the inferred property affects inductive reasoning; likewise, they do not address the possibility that different measures of similarity may underlie reasoning about different kinds of properties. However, these two simplifying assumptions, about properties and about similarity, concern issues that are central to inductive reasoning.

In an early study of the role of similarity in inductive reasoning, Rips (1975) had subjects make probability judgments about inductive statements involving pairs of animal categories. For example, given the premise that all horses on some island have a certain disease, what proportion of the mice on the island would have the disease? Another group of

---

Evan Heit and Joshua Rubinstein, Department of Psychology, University of Michigan.

This research was supported by a National Institute of Mental Health Individual Postdoctoral Fellowship to Evan Heit and by National Science Foundation Grant 91-10245 to Douglas Medin. We are grateful to Douglas Medin, Edward E. Smith, Arthur Markman, Susan Gelman, and James Hampton for comments on this research.

Correspondence concerning this article should be addressed to Evan Heit, who is now at Department of Psychology, Northwestern University, 2029 Sheridan Road, Evanston, Illinois 60208. Electronic mail may be sent to [heit@nwu.edu](mailto:heit@nwu.edu).

---

<sup>1</sup> In the terminology of Tversky (1977), anatomical features will be particularly salient when anatomical inferences are made, and behavioral features will be salient for behavioral inferences.

subjects rated the pairwise similarities between the various kinds of animals. Rips found that the inductive judgments were well predicted from the similarity judgments. However, all of the statements referred to the same property, an unnamed disease, so it is not possible to discover from this research the effect that the property itself had on reasoning.

More recent work, by Osherson et al. (1990; Osherson et al., 1991), presented a mathematical model that successfully predicted probability judgments for complex, multiple-category inductive statements. This model derived its predictions from similarity measures obtained from other subjects, again pointing to the role of overall similarity in inductive reasoning. As in Rips (1975), the stimuli for these studies were inductive statements about biological properties of animals. Again, these studies obscured the possibility that different properties might have different effects on inductive reasoning. In one set of studies, Osherson et al. (1990) used unfamiliar properties concerning anatomy or body chemistry, such as "has sesamoid bones," and these different properties were not analyzed separately. In another set of studies Osherson et al. (1991) had subjects make inductive inferences about an imaginary biological property labeled only *P*. Therefore, the conclusions of these studies are limited by the homogeneous set of properties that was used.<sup>2</sup>

### Research on Property Effects

Meanwhile, other research has focused on differences in reasoning about different kinds of properties. The studies in this second group mainly addressed the issue of which properties lead to strong inductive inferences. In Goodman's (1955) terms, properties differ on the basis of whether they are *projectible*. Projectible properties are those that people tend to make inductive inferences about. In contrast, people tend not to make inferences about nonprojectible properties, such as properties that are idiosyncratic, transient, or ill formed. (Goodman's classic example was that *green* is considered projectible, but *grue*, defined as green before Time *t* and blue after Time *t*, is not projectible.)

Nisbett et al. (1983, Experiment 1) tested subjects on inductive statements of the following form: Given that you observe that one member of Category A has Property P, what percentage of the members of Category A have Property P? Nisbett et al. found property effects. For example, given that one member of a certain tribe of people is obese, adult subjects estimated that less than 40% of the members of the tribe are obese. But given that one tribe member has a certain color of skin, subjects concluded that over 90% of the other tribe members would have the same property. Nisbett et al. showed that people treat homogeneous properties (e.g., skin color) as more projectible than properties that would be more heterogeneous (e.g., obesity) for a particular category.

In addition, recent developmental research has shown that children distinguish between kinds of properties for induction. Gelman (1988, Experiment 1) showed property effects by varying the similarity between the premise and conclusion categories as well as by varying the generalizability of the properties. Generalizable properties referred to permanent,

intrinsic characteristics such as composition (e.g., "made out of cellulose"). In addition, a set of temporary or idiosyncratic properties was intended to be nongeneralizable (e.g., "has gum stuck on the bottom"). Young children evaluated inductive inferences for both kinds of properties. Gelman found that the probability of inducing a generalizable property was correlated with the conceptual similarity between the premise and conclusion categories, but responses were at a chance level for the nongeneralizable properties. Although this study showed that children treat some properties as projectible and some properties as nonprojectible, it did not reveal different patterns for different generalizable properties.

Gelman and Markman (1986, Experiment 3) also found property effects. For biological properties, the subjects favored inferences between biologically similar animals (e.g., two snakes) over inferences between two perceptually similar, biologically dissimilar animals (e.g., a snake and a worm). For perceptual properties (e.g., softness), the children showed no preference overall between biologically matched animals and perceptually matched animals. However, this study did provide suggestive evidence that *some* children infer perceptual properties on the basis of perceptual similarity in that a few subjects (4 of 20) did consistently choose perceptually matched animals at a level greater than chance.

Springer (1992) obtained property effects that were much like Gelman and Markman's (1986) results. He compared inferences between biologically similar pairs of animals (such as a mother and a child) to inferences between pairs of animals that are perceptually similar and socially related (e.g., playmates of different species). For biological properties (e.g., "hairy ears"), Springer found that young children were more likely to make inferences between biologically matched animals than between the perceptually and socially related pairs. But for idiosyncratic properties (e.g., "very dirty from playing in mud"), subjects were equally likely to make inferences for the two pairs.

Carey (1985) compared inductive reasoning for two properties. One property was having a certain internal organ (referred to as "spleen" for children and "omentum" for adults), and the other property was having certain microscopic particles, called "golgi," inside. Carey found that subjects generally responded the same way to *spleen/omentum* and to *golgi* for a variety of animal, plant, and artifact categories. The exception was that *golgi* was more likely than *spleen/omentum* to be inferred as a property of flowers because subjects thought it was unlikely that an animal organ would also be inside a plant. Otherwise, Carey's work with induction tasks did not show property effects.<sup>3</sup>

<sup>2</sup> A recent article by Sloman (1993) proposed an alternative mathematical model, which accounts for much the same phenomena as the Osherson et al. (1990, 1991) model. The two models are alike in that they make the simplifying assumption that inferences about different properties use the same measure of similarity.

<sup>3</sup> Carey's (1985) monograph also examined children's *attributions* of several different properties, such as whether dogs sleep, have bones, and so forth. Although Carey did find property differences here, it is difficult to interpret these results for our purposes because the attribution task is so unconstrained. No premises were given to the

Together, the developmental studies suggest that for induction of biological properties, judgments for a given pair of animals are well predicted by the similarity of the animals in terms of shared biological characteristics. But for temporary or idiosyncratic nonbiological properties, the inferences appear to be largely random.<sup>4</sup> The study by Gelman and Markman (1986) leaves open the possibility that perceptual similarity has a particular influence on inductions about perceptual features.

### *Theoretical Issues*

Previous research on inductive reasoning leaves two important questions unanswered. First, is there only a single similarity measure used in inductive reasoning about properties of animals? The models of induction tested by Rips (1975) and Osherson et al. (1990; Osherson et al., 1991) relied on a single measure, such as subjects' ratings of overall similarity. Furthermore, these models do not include a mechanism for changing the similarity measure used for induction. The results of Gelman (1988) and Springer (1992) can also be explained in terms of a single similarity measure. However, much previous research on similarity judgments has shown that similarity can be evaluated in a more flexible manner, with the weights of particular features being influenced by context (Tversky, 1977; see Medin, Goldstone, & Gentner, 1993, for a review of more recent research). In this article, we investigate such dynamic use of similarity in inductive inferences.

Second, as some previous research has suggested (Goodman, 1955; Gelman, 1988; Nisbett et al., 1983), can properties simply be described as projectible or nonprojectible? We hypothesize that aside from there being good properties and bad properties for induction, what matters is the agreement between properties and similarity. For an anatomical Property P, people will make an inference between two animals (A and B) when there is anatomical similarity between Animals A and B. Likewise, they will make an inference about a behavioral property in the presence of behavioral similarity. For example, an anatomical property may be nonprojectible if Animals A and B are similar behaviorally but not anatomically. In the following experiments we focus on *when* properties are considered projectible, rather than try to distinguish between good and bad properties.

### **Experiment 1**

This experiment tested the hypothesis that there are interactive effects between the kind of match, or similarity, between a pair of animals in an inductive statement and the kind of property that is evaluated. We expected that statements with the same kind of match and property (both anatomical or both behavioral) would get higher probability ratings than statements with different kinds of match and property.

The experiment was designed with two particular methodological goals in mind. The first goal was to distinguish between the inductive support for a conclusion and the prior

belief in that conclusion. When people judge the probability that Category C has Property P, given that Animal A has Property P, they are influenced by both their prior knowledge about C as well as the additional information provided by the premise, A has P. We controlled for prior knowledge by comparing ratings for the same conclusion Category C. Say that P is an anatomical property, A is an animal that is similar to C in terms of anatomy, and B is an animal that is similar to C in terms of behavior. We compared ratings of the probability that C has Property P for two groups of subjects: those given the premise that A has P and those given the premise that B has P. The subjects surely had preconceptions about whether C has P, but this prior knowledge would affect the subjects in both conditions. The difference in responses for the two conditions should reflect only the difference in inductive support provided by the two premises A has P and B has P.

The second goal was to seek general results; therefore, we used a variety of animals and properties. Additionally, the stimuli were created from counterbalanced combinations of these animals and properties. For example, when a subject evaluated the conclusion that snakes have anatomical Property P given that lizards have P, the Property P was randomly chosen from 1 of 14 anatomical properties. This method allowed us to compare reasoning about an average anatomical property versus an average behavioral property.

### *Method*

**Subjects.** Forty-one University of Michigan undergraduates participated; they received course credit. All experiments described in this article were run according to American Psychological Association and University of Michigan ethical guidelines. No subject participated in more than one of these experiments.

**Design.** This experiment had a within-subjects, three-factor design. The two factors of major concern were kind of match between premise and conclusion category (anatomical or behavioral) and kind of property (anatomical or behavioral). The third factor was the animal category used in the conclusion of the inductive argument; seven conclusion categories were used.

**Stimuli.** We constructed seven triples of familiar animal names, shown in Table 1, according to the following criteria. In each triple, one animal was the conclusion category, and the other two animals in the triple were the premise categories. One premise category, the anatomical match, was intended to be anatomically similar to the conclusion category. These two animals were always from the same biological class (e.g., both mammals or both birds). For example, for the conclusion category whale, the anatomical match was bear. The other premise category in each triple, the behavioral match, was intended to be similar to the conclusion category in terms of behavior. Anatomical similarity was minimized by choosing the behavioral match from a different biological class than the conclusion category (e.g., if the conclusion category was a mammal, then this premise category could be a fish). The behavioral match for whale was tuna. To provide some generality, the basis of the behavioral similarity was varied. For

subjects, so they could have chosen any premise. Also, sometimes the subjects may have just retrieved known animal facts from memory rather than performed inductive inferences.

<sup>4</sup> In a recent study by Armstrong (1991), adult subjects made inferences about biological properties as well as idiosyncratic properties (e.g., "Can be taught to ring a bell.") As in the developmental studies, the adult subjects did not seem to respond systematically to the idiosyncratic properties.

Table 1  
*Mean Percentage Judgments for Inferences in Experiment 1*

Premise– conclusion category	Anatomical match		Behavioral match	
	Anatomical property	Behavioral property	Anatomical property	Behavioral property
Bear–whale	33.0	20.0		
Tuna–whale			45.7	37.0
Mouse–bat	37.9	32.6		
Sparrow–bat			33.7	44.8
Lizard–snake	68.9	69.3		
Worm–snake			49.4	58.6
Trout–shark	54.0	43.8		
Wolf–shark			17.9	20.0
Robin–hawk	68.8	52.6		
Tiger–hawk			31.2	37.6
Grasshopper– mosquito	36.6	34.9		
Vampire bat– mosquito			24.9	29.8
Ant–bee	46.7	45.1		
Hummingbird–bee			32.0	42.4

whale, bat, and snake, the behavioral match categories were intended to be similar in terms of locomotion. For shark and hawk, the behavioral match categories were intended to be similar in terms of predatory behavior. Finally, for mosquito and hummingbird, the behavioral match categories were intended to be similar in terms of food-gathering techniques.

We also created a set of 28 properties, 14 anatomical and 14 behavioral. The anatomical properties, shown in Table 2, referred to body chemistry or internal organs. The behavioral properties, also shown in Table 2, referred to acts of locomotion, predation, or feeding. (Six properties concerned locomotion, and 4 properties each related to predation and feeding.) These properties were intended to be meaningful but unfamiliar so that subjects would take them as plausibly either true or false of the animals used as stimuli. We did not assume that people's prior beliefs about the truth of these properties would be the same for different animals, and interpretation of the results does not depend on such an assumption.

Each subject was given 28 questions of the form "Given that Species A has the Property P, how likely is Species B to have the Property P?" The 28 questions were constructed randomly for each subject, with the following constraints to implement the factorial design. Each property was used for only one question. Each conclusion category was used in four questions; for two of these questions, the premise category was the corresponding anatomical match, and for the other two questions, the premise category was the behavioral match. Furthermore, each premise category appeared once with an anatomical property and once with a behavioral property. The final constraint was that behavioral properties were assigned to questions according to the kind of behavioral match that had been assigned to the conclusion animal. For example, because the basis of the behavioral match between whale and tuna was locomotion, all questions involving whales and behavioral properties used properties related to locomotive behavior rather than predation or food gathering.

Thus, the stimuli included 14 questions about anatomically matched pairs of animals, with 7 concerning anatomical properties and 7 concerning behavioral properties. Likewise, there were 14 questions about behaviorally matched pairs of animals, with 7 concerning anatomical properties and 7 concerning behavioral properties. Table 3 shows examples of these four types of questions.

**Procedure.** Subjects read the 28 questions in a booklet, with 2 or 3 questions on a page. They responded with probability judgments,

ranging from 0% to 100%, representing their degree of certainty that the conclusion category had the property in question. The experiment typically lasted 15 min.

## Results

The primary results are the responses on the four kinds of questions; they are averaged over the seven conclusion categories and shown in Figure 1. (This averaging process also implicitly averages across properties, because for each subject, properties were assigned randomly to pairs of animals. For example, the anatomical property responses for the bear–whale condition include judgments for all 14 anatomical properties, although each subject responded for only 1 property.) The kind of match between the premise and conclusion categories (anatomical or behavioral) and the kind of property (anatomical or behavioral) clearly had interactive effects. For anatomically matched pairs, subjects made stronger inferences for anatomical properties than for behavioral properties. Analogously, for behaviorally matched pairs, subjects made stronger inferences for behavioral properties than for anatomical

Table 2  
*Anatomical and Behavioral Properties for Experiment 1*

Properties
Anatomical
Its liver has two chambers that act as one.
About 70% of the optic nerve from each eye goes directly to its brain stem.
Its blood contains between 2% and 3% potassium.
Its skeletal structure contains about 58% calcium.
Its brain continues to develop during the first third of its life.
Its sense of vision is the first sense to develop.
Its sense of hearing is the last sense to develop.
Some cells in its respiratory system require carbon dioxide to function.
There are cells in its nervous system that do not require oxygen.
Its maximum heart rate is 270 beats per min.
Its outermost layer of skin is comprised of between 60 and 65% dead cells.
As it ages, its body contains more salts.
Its heart rate is not affected by nicotine.
The cartilage between its joints will grow back if torn.
Behavioral
It usually travels in a back-and-forth, or zig-zag, trajectory.
It never travels directly in the direction of the sun.
It escapes predators by random movement, so its path cannot be predicted.
It travels shorter distances in extreme heat.
After eating, it travels at speeds of twice its body length per second.
It travels in an alternating fast–slow–fast–slow pattern.
It won't eat something that it hasn't killed.
It prefers to feed at night.
It will watch its potential food for hours before attacking that food.
It won't feed if it smells any strange odors that it can't identify.
It usually gathers large amounts of food at once.
It usually defends the surrounding area near its last source of food.
Just after gathering food, it swallows 15% of the food for energy.
It will only gather food for a maximum of 2 hr at any one time.

cal properties. However, inferences seemed to be stronger overall for anatomical pairs than for behavioral pairs.

A three-way analysis of variance (ANOVA), with premise–conclusion match, property type, and conclusion category (with seven levels) as factors, supported these observations. (The cell means are shown in Table 1.) We found a reliable main effect of premise–conclusion match,  $F(1, 40) = 24.22$ ,  $p < .001$ ,  $MS_e = 1,177$ , which showed that the judgments were greater for anatomical matches than for behavioral matches. We found no main effect of property type ( $F < 1$ ), indicating that the two sets of properties did not differ in terms of how likely the subjects judged them to be true of the conclusion categories. Most important, the interaction between premise–conclusion match and property type was statistically reliable,  $F(1, 40) = 15.66$ ,  $p < .001$ ,  $MS_e = 647$ .

The conclusion category was included as a factor in the ANOVA to determine whether the central result, the interaction between premise–conclusion match and property type, generalized across the seven conclusion categories. The generality of this two-way interaction was supported because the three-way interaction, between conclusion category and the two-way interaction, was not reliable ( $F < 1$ ). There was a main effect of conclusion category,  $F(6, 240) = 24.25$ ,  $p < .001$ ,  $MS_e = 18,283$ , as well as interactions between conclusion category and premise–conclusion match,  $F(6, 240) = 10.76$ ,  $p < .001$ ,  $MS_e = 957$ , and between conclusion category and property type,  $F(6, 240) = 2.80$ ,  $p < .05$ ,  $MS_e = 498$ . These three interactions with conclusion categories were not surprising, because we had not attempted to equate the conclusion categories in terms of prior beliefs about the various properties or in terms of their degrees of similarity to various premise categories.

*Analyses by property.* What is critical about our results is that subjects showed different patterns of inference for the two sets of properties used in the experiment. We created the two

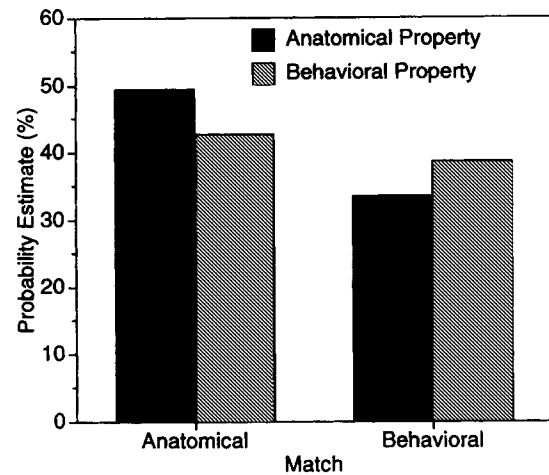


Figure 1. Mean probability judgments for Experiment 1.

sets of properties with the intention of highlighting these two different patterns of inference, rather than with the intention of somehow drawing conclusions that are general to all anatomical and all behavioral properties. Nonetheless, it is interesting to address the issue of generality across properties. We performed a two-way ANOVA with premise–conclusion match and property type as the fixed factors, and property was treated as a random factor with 28 levels. The results were equivalent to the previous analysis with subjects treated as a random factor. Again, premise–conclusion match had a reliable effect,  $F(1, 26) = 13.40$ ,  $p < .001$ ,  $MS_e = 87$ , and there was no main effect of property type,  $F(1, 26) < 1$ . The interaction between these two factors was reliable,  $F(1, 26) = 6.23$ ,  $p < .02$ ,  $MS_e = 87$ . Therefore, the results appear to be consistent across properties and not because of idiosyncratic beliefs about a few of the properties. Furthermore, this analysis suggests that our results would generalize to other properties sampled in the same manner as the properties in this experiment (Clark, 1973).<sup>5</sup>

## Discussion

The interaction found between kind of match (anatomical or behavioral) and kind of property (anatomical or behavioral) supports our claim that inductive reasoning is sensitive to the differences between the two sets of properties. Inductive inferences between anatomically matched pairs, as well as between behaviorally matched pairs, were stronger when the corresponding kind of property was inferred. These results

Table 3  
Examples of Stimulus Questions for Experiment 1

Question
Anatomical match, anatomical property Given that the species bear has the following property; Its liver has two chambers that act as one. How likely is the species whale to have the property? Its liver has two chambers that act as one.
Anatomical match, behavioral property Given that the species bear has the following property; It usually travels in a back-and-forth, or zig-zag, trajectory. How likely is the species whale to have the property? It usually travels in a back-and-forth, or zig-zag, trajectory.
Behavioral match, anatomical property Given that the species tuna has the following property; Its liver has two chambers that act as one. How likely is the species whale to have the property? Its liver has two chambers that act as one.
Behavioral match, behavioral property Given that the species tuna has the following property; It usually travels in a back-and-forth, or zig-zag, trajectory. How likely is the species whale to have the property? It usually travels in a back-and-forth, or zig-zag, trajectory.

<sup>5</sup> We did not perform an analysis that treated conclusion categories as a random effect because the seven conclusion categories were specifically chosen to unconfound the usual correlation between anatomy and behavior. Whales, bats, snakes, and the other animals we used are rather unique; it is not clear whether there are many other animals that we could have chosen. In the terminology of Clark (1973), our sample of seven conclusion categories “depleted the population” of animal categories that would allow us to unconfound anatomy and behavior. Thus, there is no need to try to generalize to a larger population.

Table 4  
*Mean Percentage Judgments in Experiment 2*

Premise– conclusion category	Anatomical match		Behavioral match	
	Anatomical property	Behavioral property	Anatomical property	Behavioral property
Rabbit–whale	31.6	30.9		
Tuna–whale			51.2	46.5
Giraffe–bat	27.4	22.6		
Sparrow–bat			41.8	55.6
Turtle–snake	46.3	46.6		
Worm–snake			55.6	62.5
Goldfish–shark	66.2	48.6		
Wolf–shark			23.7	36.0
Chicken–hawk	57.5	42.2		
Tiger–hawk			40.1	49.0
Ladybug–mosquito	42.0	35.1		
Vampire bat– mosquito			34.0	54.9
Praying mantis–bee	55.1	41.3		
Hummingbird–bee			33.8	33.5

cannot be explained by any account of inductive reasoning that does not consider the property that is being inferred. Also, because there was no difference overall in ratings about anatomical properties and ratings of behavioral properties, the results cannot be explained in terms of differing projectibility of properties in the two sets.

During debriefings, a few subjects stated that they had strong prior beliefs about a few combinations of properties and animals (e.g., bats and vision or whales and body salt). The counterbalancing in the experiment would prevent a small number of prior associations from interfering with the overall results. It is clear from Table 1 and Figure 1 that probability judgments about the conclusions were influenced substantially by the premises. Nonetheless, we used a somewhat different set of properties for the next experiment, and the problematic properties that were noted by subjects were replaced.

### Experiment 2

Experiment 1 demonstrated a difference in responding to the two kinds of properties for the two kinds of matches between animals; this suggests that inferences about anatomical properties and inferences about behavioral properties involve different similarity measures. Figure 1 and Table 1 show that anatomically matched pairs of animals generally led to stronger inferences than behaviorally matched pairs of animals, even for behavioral properties. An even more compelling result would be a true crossover interaction, which demonstrates that different properties lead to different orderings of the two kinds of pairs of animals. Another look at the anatomically matched pairs in Table 1 suggests why the behavioral inferences were so strong for these pairs of animals. It seems that the anatomically matched pairs (e.g., lizard and snake and shark and trout) share many behavioral characteristics as well. Therefore, in Experiment 2, we replaced the seven animals used as anatomical matches in the premises. The new anatomically matched premise categories were expected to be less similar to the conclusion categories in terms of behavior. However, these new anatomically matched animals were

chosen according to the same criterion as in Experiment 1 (i.e., anatomical pairs of animals had to be in the same biological class, such as mammal or bird).

Thus, Experiment 2 had two purposes. First, it was intended to replicate the result of Experiment 1 by showing that people respond differently to the two kinds of properties. Second, it was intended to strengthen this result by showing that inferences about different properties could lead to reorderings of judgments about the pairs of animals.

### Method

*Subjects.* Thirty-six subjects participated; they were recruited in the same manner as in Experiment 1.

*Stimuli and procedure.* Experiment 2 was identical to Experiment 1 in terms of stimuli and procedure, with two exceptions. First, the anatomically matched premise categories were replaced, as shown in Table 4. Second, 5 of the 28 properties were replaced, as shown in Table 5.

### Results

The responses on the four kinds of questions, averaged over the seven conclusion categories, are shown in Figure 2. These results show a crossover interaction. Now, for anatomical properties, subjects made stronger inferences between anatomically matched animals than between behaviorally matched animals, but for behavioral properties, inferences were stronger for behavioral matches than for anatomical matches.

A three-way ANOVA supported these observations. (The cell means are shown in Table 4.) The main effects of match and property type were not reliable (both  $F_s < 1$ ). However, the interaction between match and property type was clearly reliable,  $F(1, 35) = 20.88, p < .001, MS_e = 839$ .

Table 5  
*Properties Replaced From Experiment 1 to Experiment 2*

Properties
Anatomical
Experiment 1
Its sense of vision is the first sense to develop.
Experiment 2
Its sense of touch is the first sense to develop.
Experiment 1
Its maximum heart rate is 270 beats/min.
Experiment 2
All of its cells contain small amounts of zinc.
Experiment 1
As it ages, its body contains more salts.
Experiment 2
As it ages, its brain contains more serotonin.
Behavioral
Experiment 1
It never travels directly in the direction of the sun.
Experiment 2
It frequently travels for hours without stopping.
Experiment 1
After eating, it travels at speeds of twice its body length per second.
Experiment 2
It tends to move in circles during colder weather.

The three-way interaction, between conclusion category and the critical two-way interaction as previously described, was just marginally reliable,  $F(6, 210) = 2.06, p = .06, MS_e = 654$ . Taken with the lack of a three-way interaction in Experiment 1, we hold to the conclusion that our central result, the Match  $\times$  Property interaction, generalizes across the conclusion categories. Furthermore, inspection of Table 4 did not reveal an interpretable, or statistically reliable, pattern to the contrary. The marginal three-way interaction in Experiment 2 could indicate simply that the size of the interaction varies slightly over conclusion categories.

In addition, there was a reliable main effect of conclusion category,  $F(6, 210) = 4.92, p < .001, MS_e = 809$ , and a reliable interaction between conclusion category and match,  $F(6, 210) = 10.42, p < .001, MS_e = 1,176$ . Finally, the interaction between conclusion category and property was not reliable,  $F(6, 210) = 1.37, ns, MS_e = 689$ .

*Analyses by property.* In a secondary analysis, with properties treated as a random factor, and responses averaged across subjects, the results were comparable. This two-way ANOVA (premise–conclusion match and property type) showed no main effect of premise–conclusion match and no main effect of property type; for each fixed factor,  $F(1, 26) < 1$ . The interaction between these two factors was again reliable,  $F(1, 26) = 8.72, p < .01, MS_e = 102$ . This analysis indicates that our results would generalize to other anatomical and behavioral properties sampled in the same manner as in Experiment 2.

### Discussion

Experiment 2 replicated the finding in Experiment 1 that people make stronger inductive inferences when the kind of match between the premise and the conclusion categories corresponds to the kind of property. Furthermore, Figure 2 shows that on the average, subjects had different orderings for which premise led to stronger inferences about a given conclusion category. That is, for anatomical properties, people preferred anatomically matched premises and conclusions, but for behavioral properties, people preferred behaviorally matched pairs.

These results logically eliminate the possibility that people have a single similarity measure supporting inductive reasoning about animals. For behavioral properties, the subjects judged the inductive arguments between two animals in the same biological class to be *weaker* than the arguments involving pairs of animals from different biological classes. This reordering suggests that when subjects made behavioral inferences, they relied on a different measure of similarity than when they made anatomical inferences. It is logically impossible for *any* single measure of similarity to correlate with the two orderings of animal pairs obtained for anatomical and behavioral properties.

Furthermore, the crossover pattern of results cannot be explained in terms of simple differences between the two sets of premise categories. Previous research by Rips (1975) and Osherson et al. (1990; Osherson et al., 1991) has shown that people are more likely to make inferences when the category in the premise of an inductive argument is typical of its superordinate. For example, people are more likely overall to make an inference about a property of dogs than about a property of

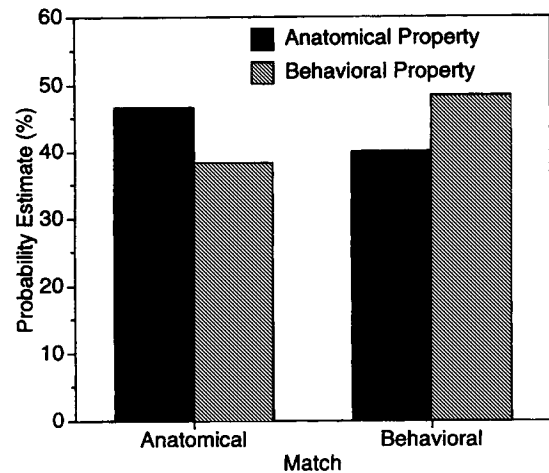


Figure 2. Mean probability judgments for Experiment 2.

squirrels because dogs are more typical than squirrels of the superordinate category, mammals (see Shipley, 1993, for further discussion of these phenomena). In the present study, any difference in typicality of premise categories would be manifested as a main effect of anatomical match versus behavioral match; however, we found no such main effect.

### Experiment 3

The purpose of this experiment was to investigate further the role of similarity in the inferences made in the previous two experiments. In particular, we hypothesized that different similarity measures were involved when subjects had made judgments about anatomical properties and behavioral properties so that subjects would focus on anatomical similarity when making anatomical inferences and behavioral similarity when making behavioral inferences. Such an effect would be analogous to compatibility effects in decision making, in which people tend to use a predicting cue that resembles the dimension to be inferred (see Slovic, Griffin, & Tversky, 1990). The two similarity measures were studied in Experiment 3 in a direct manner by asking subjects to make similarity judgments *with respect to* anatomy or behavior. We expected that people would focus on anatomical features when they judged similarity with respect to anatomy, and likewise they would focus on behavioral features when they judged behavioral similarity. Therefore, we expected that the similarity judgments with respect to anatomy or to behavior obtained in Experiment 3 would be correlated with the inductive inferences about the corresponding kinds of properties that were obtained in Experiments 1 and 2.

Furthermore, this experiment served as a manipulation check for our choices of animals in the first two experiments. Although these experiments were successful, they did rely on stimuli that were chosen on the basis of the experimenters' intuitions. The anatomical and behavioral similarity judgments were intended to provide independent evidence that the anatomically matched animal pairs were considered similar in terms of anatomy and that the behaviorally matched animal pairs were considered similar in terms of behavior.

Table 6  
*Anatomical and Behavioral Similarity Ratings  
 From Experiment 3*

Animals	Similarity	
	Anatomical	Behavioral
Whale-bear	3.29	3.10
Bat-mouse	4.99	3.46
Snake-lizard	6.47	5.96
Shark-trout	6.56	4.88
Hawk-robin	7.29	4.60
Bee-ant	5.15	4.35
Mosquito-grasshopper	4.43	3.01
Whale-rabbit	2.51	2.83
Bat-giraffe	2.03	1.64
Snake-turtle	4.07	3.14
Shark-goldfish	5.75	3.60
Hawk-chicken	6.44	3.08
Bee-praying mantis	4.43	3.19
Mosquito-ladybug	5.43	3.53
Whale-tuna	5.56	5.87
Bat-sparrow	4.81	5.17
Snake-worm	4.90	5.25
Shark-wolf	2.32	6.08
Hawk-tiger	2.29	5.72
Bee-hummingbird	3.40	6.64
Mosquito-vampire bat	3.19	7.18

## Method

**Subjects.** Seventy-two University of Michigan undergraduates participated in this study; they received either course credit or a small honorarium. Subjects judged similarity both with respect to anatomy and with respect to behavior. The subjects were assigned randomly to two groups so that half made anatomical judgments first and half made behavioral judgments first.

**Stimuli.** The instructions first explained the idea of focusing on certain characteristics when judging similarity. The example given was that when one focuses only on color, the sky and the ocean are quite similar. For the anatomical similarity judgments, the subjects were instructed to focus on anatomical and biological characteristics, such as internal organs, bones, genetics, and body chemistry. For the behavioral similarity judgments, they were instructed to focus on behavioral characteristics, such as movement, eating habits, and food-gathering or hunting techniques. We expected that these instructions would lead subjects to consider a variety of familiar animal characteristics besides the properties investigated directly in Experiments 1 and 2.

The 21 pairs of animals included the 7 anatomical pairs used in Experiment 1, the 7 anatomical pairs from Experiment 2, and the 7 behavioral pairs used in both experiments. These pairs are shown in Table 6. The stimuli were presented on paper, in one of four random orders.

**Procedure.** Subjects first rated all 21 animal pairs with respect to either anatomy or behavior. Then they were presented with the animal pairs again, in a different order, and they made the other kind of similarity rating. The ratings ranged from *not at all similar* (1) to *highly similar* (9). The procedure typically lasted 15 min.

## Results

**Manipulation checks.** The first finding was that the second set of ratings made by subjects appeared to be unaffected by making the first set of ratings, despite the change of focus.

Across the 21 animal pairs, there was a correlation of .95 between the mean anatomical similarity judgments of subjects who made these ratings first and of subjects who made these judgments second. In other words, previously made behavioral judgments did not affect the ordering of subjects' anatomical judgments. Likewise, the correlation over the 21 animal pairs for behavioral similarity judgments by the two groups of subjects was over .95. (In addition, the high level of agreement between the two groups of subjects serves as evidence of the reliability of the measures used.) Therefore, the subsequent analyses included the judgments of all subjects without regard to which kind of judgments they made first. The average judgments for each animal pair are shown in Table 6.

Next, as a manipulation check to see if subjects were truly focusing on different characteristics when they judged similarity with respect to anatomy and with respect to behavior, we compared these two kinds of judgments. Across the 21 animal pairs, the correlation between the average anatomical and behavioral judgments was only .03. Clearly, the subjects were making two distinct kinds of similarity judgments.

Finally, the similarity ratings were used to check our experimental manipulation of animal pairs. These ratings confirmed our previous intuitions. Table 7 shows the average ratings made on the anatomical pairs from Experiment 1, the anatomical pairs from Experiment 2, and the behavioral pairs used in both experiments. For the anatomical pairs from Experiments 1 and 2, the anatomical similarity ratings were greater than the behavioral similarity ratings; using a paired *t* test,  $t(71) = 3.62, p < .001$ , and  $t(71) = 7.72, p < .001$ , respectively. We chose different anatomical pairs for Experiment 2 with the intention of reducing the behavioral similarity between these animals, and Table 7 shows that we successfully reduced the behavioral similarity for anatomical pairs from Experiment 1 to Experiment 2. Additionally, the behavioral pairs received greater behavioral similarity ratings than anatomical similarity ratings,  $t(71) = 10.18, p < .001$ .

**Regression analyses.** The most important analysis was to determine the relation between the anatomical and behavioral similarity ratings from Experiment 3 and the evaluations of the inductive arguments from Experiments 1 and 2. Experiments 1 and 2 were analyzed separately because they involved different animals as well as a few different properties. Each experiment included 14 premise-conclusion pairs of animals, for which each subject judged the probabilities of both anatomical properties and behavioral properties. We used multiple regression analyses to predict these mean probability judgments, for each pair, from the two kinds of similarity judgments. Although the experimental design of Experiments 1 and 2 was focused on finding differences between the two groups of

Table 7  
*Mean Anatomical and Behavioral Similarity Ratings*

Matches/Experiment	Similarity	
	Anatomical	Behavioral
Anatomical matches/Experiment 1	5.28	4.79
Anatomical matches/Experiment 2	4.24	2.93
Behavioral matches/Experiments 1 and 2	3.33	4.96



animal pairs, the regression analyses permit us to examine differences within each group as well. For example, on the whole, the 7 behaviorally matched pairs of animals were quite similar in terms of behavior but not very similar in terms of anatomy. However, Table 6 shows that these 7 pairs of animals differed substantially from each other in terms of anatomical and behavioral similarity. Therefore, regression analyses of the judgments on the 14 individual pairs might reveal patterns not evident from the previous blocked analyses.

The resulting regression equations are shown in Table 8. For example, the best-fitting regression equation to predict the inductive judgments about anatomical properties for animal pairs in Experiment 2 is

$$\text{Probability estimate} = 7.58 * \text{Anatomical similarity} \\ + -0.60 * \text{Behavioral similarity} + 15.05, \quad (1)$$

and the regression equation for predicting judgments about behavioral properties is

$$\text{Probability estimate} = 3.86 * \text{Anatomical similarity} \\ + 3.26 * \text{Behavioral similarity} + 12.80. \quad (2)$$

In both experiments, the probability judgments for inductive arguments involving anatomical properties are well predicted by anatomical similarity between animals. The regression coefficients for anatomical similarity are both reliably greater than zero, but the coefficients for behavioral similarity are near zero. In contrast, for both experiments, the probability judgments for behavioral properties seem to have been influenced by both behavioral similarity and anatomical similarity. The regression coefficient for behavioral similarity is reliable for Experiment 2 ( $p < .05$ ) and marginally reliable for Experiment 1 ( $p < .06$ ). In addition, anatomical similarity is a reliable factor in both equations.

In summary, the regression analyses revealed that inferences about anatomical properties are related to anatomical similarity, and they revealed the unexpected finding that inferences about behavioral properties are related to both behavioral and anatomical similarity.

## Discussion

This experiment supports the claim that people can differentially judge similarity with respect to anatomy and behavior and that these two kinds of similarity have distinct contributions to inductive reasoning. What is most surprising about the results of the regression analyses is the asymmetry between inductive reasoning about the two kinds of properties. As would be expected from the ANOVAs of Experiments 1 and 2, inferences about anatomical properties were related to anatomical similarity and not to behavioral similarity. But people's inferences about behavioral properties were influenced by both behavioral similarity and anatomical similarity. Thus, when people evaluated inferences about behavioral proper-

Table 8  
*Regression Equations Predicting Mean Percentage Judgments for Inferences About Anatomical and Behavioral Properties*

Property	Coefficient			$R^2$
	Anatomical similarity	Behavioral similarity	Constant	
Experiment 1				
Anatomical				
$M$	8.62	0.16	0.35	.79***
CI	5.57, 11.68	-3.57, 3.89		
Behavioral				
$M$	6.98	4.08	-11.78	.64**
CI	3.42, 10.54	-0.27, 8.43		
Experiment 2				
Anatomical				
$M$	7.58	-0.60	15.05	.80***
CI	5.05, 10.12	-2.81, 1.61		
Behavioral				
$M$	3.86	3.26	12.80	.46*
CI	0.22, 7.50	0.10, 6.43		

Note. Values in parentheses show the 95% confidence intervals (CI) for regression coefficients.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

ties, they did not simply judge similarity by considering other behavioral properties. One possible interpretation of the asymmetry is that people believe in an asymmetric causal relationship between anatomy and behavior (cf. Keil, 1989; Medin & Ortony, 1989). Both anatomical and behavioral characteristics affect an animal's behavior, so both kinds of similarity are relevant for inferring behavior. However, people do not believe that behavior has consequences for anatomy, so behavioral similarity does not influence inferences about anatomy.

Another implication of this asymmetric result is that it demonstrates the importance of biological categories. Anatomical similarity, referring to the insides of animals, is a better indicator of biological category than is behavioral similarity. As illustrated by our stimuli, behaviorally similar animals need not come from the same biological category (consider snakes and worms or hawks and tigers). Even for inferences about behavior, we found that anatomical similarity had a significant influence. Therefore, our studies maintain a continuity with previous research (e.g., Gelman and Markman, 1986), which demonstrated the importance of biological categories for inductive reasoning. Furthermore, our results suggest that reasoning about animals in terms of anatomy and in terms of behavior are not simply two encapsulated perspectives because inferences about behavior were affected by nonbehavioral anatomical information.

An additional finding of the regression analyses in Experiment 3 is that the similarity ratings were better predictors of the anatomical property inferences than of the behavioral property inferences, as shown by the squared multiple correlation values in Table 8.<sup>6</sup> This result suggests that to the extent

<sup>6</sup> The comparison of squared multiple correlation values for anatomical inferences and behavioral inferences is merely suggestive because the behavioral probability judgments in the two experiments had somewhat lower ranges and variances than the anatomical

that factors other than similarity play a role in inductive reasoning, these other factors may have a greater role in the inferences about behavioral properties. Another interpretation of this difference between the two sets of regression equations is that our measure of anatomical similarity was better than our measure of behavioral similarity. The behavioral properties were split into three groups: locomotion, eating, and predation, but we tried to predict these three kinds of inferences with an undifferentiated behavioral similarity measure in terms of all three kinds of behavior. We would expect that even more focused similarity ratings, with respect to only locomotion, eating, or predation, would lead to a better account of behavioral inferences.

Finally, it is striking that we found a zero correlation between anatomical and behavioral similarity ratings. This finding surely validates our manipulation of the similarity rating task, but we do not want to claim that over a larger group of animals that there would be no correlation between anatomical and behavioral characteristics. Our animal pairs were especially chosen to unconfound the correlation between anatomy and behavior. We would expect a positive correlation between the two measures over a wider set of animal pairs, but the zero correlation that we obtained suggests that the two measures are indeed distinct.

## General Discussion

### *How Many Similarity Measures?*

We have shown different patterns of inductive reasoning for two kinds of properties, in terms of two similarity measures: anatomical and behavioral. The simplifying assumption of much previous research, that only a single similarity measure influences inferences about different properties, cannot account for our results. In addition to the evidence from our experiments, we see other reasons to reject the idea of a single similarity measure for induction. First, research by Barsalou (1989) has suggested that when people make judgments about a category, they retrieve a largely context-dependent subset of its features rather than all of its features. (Goodman, 1972, has made a related point, that categories may be considered to have a potentially infinite number of features.) Therefore, it is more plausible that similarity will depend on different features in different contexts and for different goals rather than similarity being a strictly fixed measure of feature matches. Second, we doubt that people's inductive inferences outside of the psychology laboratory would be successful if people were restricted to using the same similarity measure for inferring different properties. A more flexible process of induction, in which the kind of property being inferred affects the similarity measure, would likely lead to more accurate inferences.

Then, are there just two similarity measures for reasoning about animal categories, anatomical and behavioral similarity, or do people use more than two kinds of similarity? One approach to find more similarity measures would be to try to

further differentiate the two measures that we have used. For example, implicit in the design of Experiments 1 and 2 is the idea that behavioral similarity can be based on characteristics of locomotion, eating, or predation. For instance, people made strong inferences about properties relating to locomotion for pairs of animals that were similar in terms of locomotion. Behavioral properties related to eating probably would not show the same pattern of inferences as behavioral properties related to locomotion. We find it plausible that people use somewhat different measures of behavioral similarity when they make inferences about different behavioral properties. However, our data do not address this issue directly because the three different groups of behavioral properties were evaluated for three different groups of animals (e.g., eating properties were not evaluated for animals similar in terms of locomotion).

### *Similarity and Theories*

There is an alternative, however, to looking for a fixed set of similarity measures used in inductive reasoning. Instead, prior knowledge could be used dynamically to focus on certain features when similarity is evaluated. In this conception, inductive reasoning is an active process in which people identify the features in the premise and conclusion categories that are relevant to the property being inferred. The more that is known about the property, the more constrained this set of features will be. In our experiments with fairly unfamiliar properties, the subjects knew just enough about the properties to treat the anatomical and behavioral properties differently.

Murphy and Medin (1985) have also proposed that people use background knowledge, or theories, to perform induction. Applying their approach to our experiments, one might argue that the subjects were bringing to bear their background knowledge about the causal relations between various characteristics of animals. For example, subjects responded in terms of their theories of how a behavioral property might be caused by, or explained in terms of, other anatomical and behavioral properties. This theory-based approach to research on inductive reasoning seems compelling, but it leaves two open questions. First, what is the content of people's theories? Second, how do people use these theories for reasoning?

The first question has been a prominent topic of study (e.g., Carey, 1985; Keil, 1989), but the second question has not been researched as thoroughly. Although there are successful psychological models for similarity-based reasoning (e.g. Osherson, et al., 1990, 1991), there do not appear to be any successful models of the kind of theory-based reasoning described by Murphy and Medin (1985). We suggest that these two kinds of reasoning would work well together.<sup>7</sup> What is lacking from

probability judgments. Thus, the coefficient of determination values may have been larger for the anatomic inferences because there was more systematic variation to predict with regression analyses.

<sup>7</sup> Although we believe that similarity has pervasive effects on inductive reasoning, we do not mean to claim that *every* instance of inductive reasoning depends on similarity. Consider this example from Smith, Shafir, and Osherson (in press). People say that the inductive argument (a) A poodle can bite through wire, therefore a German shepherd can bite through wire, is stronger than (b) A doberman can bite through wire, therefore a German shepherd can bite through wire. Despite the greater similarity between dobermans and German

models of similarity-based reasoning is the specification of which features will be considered in the similarity measure. The theory-based approach provides the intuition that we know that some features are causally related to others, but this approach leaves out a description of how people would use this feature knowledge for reasoning.

Our proposal is that in inductive reasoning, people's theories constrain their similarity measures. To paraphrase Keil (1989), similarity-based reasoning takes over where theory-based reasoning leaves off. When people make an inference about Property P, they first consider, on the basis of their causal knowledge about P and other properties, which properties are relevant to P. Then they reason on the basis of similarity with respect to those properties that are deemed relevant. Given the arguments of Goodman (1972) about potentially infinite sets of features, it is hard to imagine that similarity-based inductive inference could take place at all without an initial step of constraining the features that are considered. In other words, all inductive reasoning that is based on similarity must depend on some assumptions about what is relevant.<sup>8</sup>

### *Relations to Categorization*

Our results have implications beyond the canonical inductive reasoning task, in premise-conclusion form, investigated here. More generally, psychological accounts of how people use similarity for reasoning must consider what they are reasoning about. For example, when people categorize new objects, situations, or people, they respond to the similarity between the new items and old examples in memory (e.g., Heit, 1992). The memorized examples serve in effect as premises for inductive inferences about the new stimuli. Consider the tasks of categorizing people as good or bad tennis players, good or bad conversationalists, and liberals or conservatives. Although similarity to known tennis players, conversationalists, or liberals can play a role in these three different kinds of categorization, clearly three different measures of similarity would be needed. The features relevant to categorizing someone as a good tennis player are not the same as the features relevant to categorizing someone as a good conversationalist (Heit, 1993).

Current models of categorization (e.g., Anderson, 1991; Medin & Schaffer, 1978; Nosofsky, 1986) do not assume that a single measure of similarity underlies all categorizing. Instead, these models have a flexible conception of similarity; they treat the relative weightings of various features as free parameters so that different measures of similarity could be implemented. Yet, like the induction research focusing on single properties, many individual studies of categorization have examined only one kind of category. However, some recent research has shown that people use different similarity measures for different categories. For example, in a study of categorization of geometric figures, Nosofsky showed that the underlying similar-

ity measures used by subjects were influenced in a systematic way by the structure of the categories that were learned. In a domain closer to the present research, Jones, Smith, and Landau (1991) found that children categorized a set of artifact stimuli on the basis of shape, but they categorized a set of animal stimuli using similarity with respect to both shape and texture.

We propose that future models of induction and categorization must develop a better account of this fundamental issue of how different measures of similarity are used. In particular, future research needs to investigate the processes by which theoretical knowledge, about which features are relevant to making an inference, affects similarity-based reasoning.

<sup>8</sup> Some of these assumptions might actually be innate constraints, as suggested by Quine (1977).

### References

- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409-429.
- Armstrong, S. L. (1991, November). *Category and property specificity in category-based induction*. Paper presented at the meeting of the Psychonomic Society, San Francisco.
- Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76-121). Cambridge, England: Cambridge University Press.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, 20, 65-95.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183-209.
- Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.
- Goodman, N. (1972). *Problems and projects*. New York: Bobbs-Merrill.
- Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, 24, 341-380.
- Heit, E. (1993, November). *Comparing models of prior knowledge effects on category learning*. Paper presented at the meeting of the Psychonomic Society, Washington, DC.
- Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, 62, 499-516.
- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254-278.
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge, England: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90, 339-363.

shepherds, people favor (a) because German shepherds are much stronger than poodles. The crucial characteristic of this counter example seems to be that the inferred property mainly depends on a single magnitude scale, strength.

- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185-200.
- Osherson, D. N., Stern, J., Wilkie, O., Stob, M., & Smith, E. E. (1991). Default probability. *Cognitive Science*, 15, 251-269.
- Quine, W. V. O. (1977). Natural kinds. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 155-175). Ithaca, NY: Cornell University Press.
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665-681.
- Shipley, E. F. (1993). Categories, hierarchies, and induction. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 265-301). San Diego, CA: Academic Press.
- Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, 25, 231-280.
- Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. In R. M. Hogarth (Ed.), *Insights in decision making* (pp. 5-27). Chicago: University of Chicago Press.
- Smith, E. E., Shafir, E., & Osherson, D. (in press). Similarity, plausibility, and judgments of probability. *Cognition*.
- Springer, K. (1992). Children's awareness of the biological implications of kinship. *Child Development*, 63, 950-959.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

Received October 26, 1992

Revision received May 3, 1993

Accepted May 10, 1993 ■

## Low Publication Prices for APA Members and Affiliates

**Keeping You Up-to-Date:** All APA members (Fellows; Members; Associates, and Student Affiliates) receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*.

High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they can subscribe to the *American Psychologist* at a significantly reduced rate.

In addition, all members and affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential Resources:** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the APA*, the *Master Lectures*, and *Journals in Psychology: A Resource Listing for Authors*.

**Other Benefits of Membership:** Membership in APA also provides eligibility for low-cost insurance plans covering life, income protection, office overhead, accident protection, health care, hospital indemnity, professional liability, research/academic professional liability, student/school liability, and student health.

**For more information,** write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242, USA