
Volume 63, Issue 1, July 2010
ISSN 0749-596X
ELSEVIER
Journal of
Memory and
Language
Editor
Victor Ferreira
Associate Editors
Sally Andrews
Simon Farrell
Robert Greene
James Magnuson
Editorial Board
David Balota
Kathryn Bock
Julie Boland
Gordon Brown
Jenn-Yeu Chen
Charles Clifton
Albert Costa
Amy Criss
Gary Dell
Kara Federmeier
Fernanda Ferreira
Cindy Fisher
Ken Forster
David Gallo
Alan Garnham
Richard Gerrig
Zenzi Griffin
Rob Hartsuiker
Evan Heit
Larry Jacoby
Debra Jared
Sachiko Kinoshita
Judy Kroll
Stephen Lupker
Michael Masson
Sven Mattys
Brian McElree
James McQueen
Alissa Melinger
Antje Meyer
Neil Mulligan
Gregory Murphy
Ian Neath
James Neely
Klaus Oberauer
Martin Pickering
Keith Rayner
Henry Roediger
Arty Samuel
Patrick Sturt
Marcus Taft
Rebecca Treiman
John Trueswell
Nash Unsworth
Jos van Berkum
Mieke Verfaellie
Duane Watson
Andrew Yonelinas
Johannes Ziegler
Available online at www.sciencedirect.com
ScienceDirect

Journal of Memory and Language 63 (2010) 1–17

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Category vs. object knowledge in category-based induction

Gregory L. Murphy [a*], Brian H. Ross [b]

[a] Department of Psychology, New York University, United States
[b] Beckman Institute and Department of Psychology, University of Illinois at Urbana-Champaign, United States

A R T I C L E I N F O

Article history:
Received 4 September 2009
revision received 9 December 2009
Available online 27 March 2010

Keywords:
Concepts
Induction
Category-based induction

A B S T R A C T

In one form of category-based induction, people make predictions about unknown properties of objects. There is a tension between predictions made based on the object's specific features (e.g., objects above a certain size tend not to fly) and those made by reference to category-level knowledge (e.g., birds fly). Seven experiments with artificial categories investigated these two sources of induction by looking at whether people used information about correlated features *within* categories, suggesting that they focused on feature–feature relations rather than summary categorical information. The results showed that people relied heavily on such correlations, even when there was no reason to think that the correlations exist in the population. The results suggested that people's use of this strategy is largely unreflective, rather than strategically chosen. These findings have important implications for models of category-based induction, which generally ignore feature–feature relations.

Introduction

Induction of properties is widely recognized as a critical function of categories. Once one has identified something as a tomato, one feels fairly safe in eating it, even without prior experience of that tomato. One also feels confident in inferring that it will not serve well as a hockey puck or seat. In these ways, the category of tomatoes spares us much trouble and makes our lives more efficient (Smith & Medin, 1981). However, it is not always clear how much category-level knowledge is involved in such inferences. For example, although tomatoes are generally not poisonous and therefore are safe to eat, some tomatoes are ripe and blemish-free, whereas others might be unripe or putrid. If you pick up a tomato, its softness makes it immediately apparent that use as a hockey puck would be ill-advised, even if you do not think about the category of tomatoes as a whole. The details of any particular exemplar, when they are known, play a role in guiding our behavior towards that object in addition to knowledge of the object's category.

In most prior research on category-based induction, only categorical knowledge has been tested. For example, in Rips's (1975) classic induction task, people are told that one or more categories of animals have a disease. They are then asked to judge the proportion of animals in another category that have the disease. The disease is ascribed to the entire given category, and no information about specific exemplars is given or asked about. Similarly, in the influential work of Osherson, Smith, Wilkie, López, and Shafir (1990), people are told categorical statements such as "Penguins have sesamoid bones" and then judge whether another category also has sesamoid bones. In this paradigm, then, only *category-level* knowledge can be used, because the details of a particular penguin or other birds are simply not given.

In everyday life, however, we often make inductions about specific exemplars, such as deciding what to feed a neighbor's cat or whether a particular oak is strong. In such cases, one has both summary category-level knowledge, such as cats liking fish and oaks being very sturdy, and

* Corresponding author. Address: Department of Psychology, New York University, 6 Washington Place, 8th floor, New York, NY 10003, United States.
E-mail address: gregory.murphy@nyu.edu (G.L. Murphy).

0749-596X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2009.12.002

specific exemplar-level knowledge, such as the cat's behavior and the oak's appearance. This specific information could lead to induction through feature–feature links, such as observing that an oak's missing branches suggest that it is old and therefore not strong. Furthermore, we may use memories of specific exemplars to make our induction, such as an old oak that fell down after a lightning strike. The goal of the present research is to widen the scope of category-based induction research to discover the relative roles of object-specific and category-level information in induction.

A well-known paradigm in conceptual development pioneered by Gelman and Markman (1986) does address induction to individual exemplars. In this paradigm, children are given information about two objects and then are asked to draw an induction to a third object, which might look like one of them but share a category label with the other. Although this is an induction about a specific object, there is no category-level information in this task. Children are only told about the properties of individual objects—no information is given about the properties of the category as a whole. That is, rather than "Dogs have blicket inside," children might be told, "This dog has blicket inside." Therefore, this paradigm does not contrast category-level vs. specific features. However, it does seem relevant that, in addition to category membership, the similarity between the given and target objects influences induction (e.g., Gelman & Markman, 1987, p. 1537), as object similarity is exemplar-level of information.

Category-based inferences

We investigated the roles of feature–feature vs. summary category-level knowledge through the issue of *feature independence*. To explain this, consider Anderson's (1991) groundbreaking model of how people infer features from categories when categorization is uncertain. (We have abstracted the induction component from the rest of Anderson's model and applied it to natural categories, rather than the categories that his model creates. We are not proposing our work as a test of Anderson's whole model, but rather as a test of the general assumptions that underlie his account of induction.) Suppose you hear an animal growl at night while you are camping and try to predict whether it will try to break into your tent. Anderson suggested that this process had two components. In one, you identify the likelihood of this animal being in the different categories you know, based on the feature (growling) you observed. For example, perhaps this growling animal is a bear; very possibly a dog; possibly a raccoon; unlikely to be a moose; and so on. Each of these categories has some probability of classifying this object. In the second component, for each category, you estimate the probability that its members would try to break into your tent: bears, 50%; dogs, 20%; raccoons, 10%; and so on. The probability of each category is multiplied by its likelihood to try to break into your tent, and the sum of these products would form your estimate that the growling animal will break into your tent.

One consequence of such an account is that the to-be-predicted feature (tent invasion) is treated as independent of the given features (growling) within each category. Perhaps bears are likely to growl and to try to get inside your tent, but those are category-level attributes. The assumption is that, within the category of bears, those that growl are no more likely to break into your tent than those that do not. We call this *the independence assumption*. (It is important to emphasize that the independence assumption applies only *within* a category. Everyone agrees that features are not independent across categories; Rosch, 1977.) Anderson (1991, p. 411) argued that feature independence is likely true when the categories are biological species, because "the display of features within a freely interbreeding species is largely independent." Whether this assumption holds widely is something we will discuss in the General Discussion. However, whether it is true in the world or not, it may well be a simplifying assumption that people use when making inductions. Furthermore, most models of category-based induction seem to share this assumption of independence, insofar as they provide no mechanism for dealing with correlated features (see General Discussion for specific examples).

The independence assumption also has the important theoretical interpretation that predictions are performed at the category level, even if the person knows features of the specific exemplar. That is, if bears are tent-enterers, then people will predict that this animal will enter the tent to the degree they think it is a bear. Details of specific "bears I have known" are not retrieved to compute the induction. The assumption allows the inducer to rely on categorical inferences and knowledge to draw this induction. Perhaps we have no specific knowledge of whether raccoons will enter tents, but by reasoning that most raccoons are afraid of people and therefore will stay out, we can give an answer to this question without worrying about specific growling vs. nongrowling raccoons.

Taking a step back, it is worth noting that almost all research on categorization has emphasized that people's reliance on categories is a central part of intelligent behavior, because it permits us to generalize our knowledge to new examples. Such statements (Anderson, 1991; Murphy, 2002; Smith & Medin, 1981) suggest that it is category-level featural knowledge that is providing the basis for inductions. This category-level knowledge is the sort of knowledge that could be described as information about the category as a whole, rather than individuals: Dogs bark; most birds fly; 10% of fish are edible. If people are paying attention to feature–feature correlations in making these inductions (raccoons that growl will enter tents, but those that are silent probably would not), then summary category-level knowledge is a less critical aspect of induction. So far as we know, Anderson's proposal is the *only* proposal in the field that addresses induction to specific exemplars, which is why it is our starting point for discussion. But this also reinforces the fact that most theories have emphasized category-level knowledge. To the degree that people try to use the visible features of an item to directly infer other features, they are not using category knowledge to generalize.

Our experiments test the independence assumption. If people use within-category correlations to make inductions, this will show that exemplar-level knowledge is important even when category-level knowledge can provide an answer. An earlier experiment (Murphy & Ross, 1994, Experiment 8) suggested that people can be sensitive to feature correlations in making inductions. However, in that study, the correlation only emphasized the feature that would be induced by the category base-rate, and so there was no clear way to distinguish category-level from exemplar-level information. The present experiments will contrast these two factors and will investigate why people might use exemplar-level information.

Why and how

Why might people use feature–feature correlations to draw inductions, contrary to the independence assumption? If this knowledge is more reliable or specific than category-level knowledge, it obviously would be beneficial to use it. For example, if cats have many different colors, but six-toed cats are predominantly calico, then in making an inference about a cat's color one would be well advised to discover how many toes it has. In other cases, people may not have encoded the target feature for different categories, and so category-level information may not be available. If you do not know whether raccoons or dogs will break into inhabited tents, you may have to draw your inference from whatever features you can presently observe in the animal in question, such as its present location and apparent aggressiveness.

How would such inductions be computed? One route is through causal or other knowledge that links one feature to another. In category learning tasks, people notice and take advantage of knowledge that links a category's features, even though the task does not require it (Kaplan & Murphy, 2000; Murphy & Wisniewski, 1989). It seems likely that people would use such knowledge in an induction task that specifically asked about the relation between two features (e.g., predicting that an animal that lives near water might eat fish). Proffitt, Coley, and Medin (2000) found that tree experts used their knowledge of ecology to help answer induction questions, although they were not asked about a particular exemplar but instead judged how widespread a disease would be. Nonetheless, their results suggest that people may use knowledge of feature–feature relations when it is available, and such knowledge might pre-empt category-level information.

Another way to compute such inductions is through empirical evidence of feature correlations. Imagine that you know that some animal has thick fur and growls, and you want to know where it lives. Rather than access category-level information about the habitats of wolverines, bison, and lions, you might think of specific animals you remember that have such features (e.g., Heit & Barsalou, 1996). For example, if you can think of 25 growling thick-furred animals you have encountered, with 20 of them living in forests, and 5 of them living in your crazy uncle's basement, you will probably infer that the new animal lives in a forest. This answer might well correspond to the answer you would have gotten if you had consulted categories like wolverines and bison. But because it depends on the particular facts known about specific exemplars, the two answers could deviate. For example, imagine that you believe that most raccoons live in suburban locations and live on human garbage. However, if you happen to have seen many growling, thick-furred raccoons in forests, then your inference may be based on these exemplars rather than your more abstract category-level belief. You simply may not use your category-level belief to make the induction, because it did not include information about thick-furred, growling raccoons.

One way to calculate such inductions is to look for feature *conjunctions*. One would search memory for objects with the given features and then tally up the target features. For example, in one's memory, one looks for raccoon exemplars with thick fur + growling and counts up the habitats stored with each exemplar in order to discover which is most frequent. If there are more thick fur + growling + forest instances retrieved than there are thick fur + growling + uncle's basement instances, then one would predict that this new item would live in the forest. A second way to compute correlations is simply to ignore category membership completely and count up the conjunction of the given and target features. Because our experiments do not distinguish these two ways of using feature–feature information (see Hayes, Ruthven, & Newell, 2007), we will lump them together as *the conjunction strategy*.

When should the conjunction strategy be used? In some circumstances, it seems perfectly reasonable, such as when category-level information is not available. However, when categories do contain useful information, there is a tension between using summary category-level information and restricting one's induction to objects that contain the given feature. In particular, if the given and target features are not correlated, then the conjunction strategy is suboptimal. If the features are not truly correlated, then the best that the conjunction strategy can do is to provide the same answer that would have been obtained by ignoring the feature relations and simply using the category base-rate. So, if a cat's color is unrelated to its toes, then considering the toes of cats you have known cannot increase your inductive accuracy. But it could easily make it worse, due to sampling error. If you have a small or biased sample of six-toed cats, growling animals, and so on, then using the conjunction strategy will access only the small number of items that you know with the given feature. That small number will generally be less accurate than if you had used your knowledge of all category members. Also, using conjunctions would be most useful when the correlation between the features is high. When the correlation is low, the answer will only deviate slightly from the category-level answer, and so any improvement in predictive power may be overwhelmed by sampling error.

On the other hand, category-level information can also be unreliable. Perhaps one has only a vague memory that a lot of raccoons carry rabies, based on a forgotten conversation; perhaps one's social stereotype is based on biased information. The question of when category-level and feature–feature information should be used depends on their relative validity. Most models of category-based induction seem to assume that the category-level information should outweigh specific featural information, since they do not use the latter at all. It will be interesting to discover whether people share this assumption.

If we find that people use a conjunction strategy, it will be important to discover whether they are sensitive to the situations when it could vs. could not improve their inductions. Perhaps they will use such a strategy when the given and target feature are correlated but use category-level information when the two are not correlated. This would suggest that their use of feature conjunctions is based on well-founded statistical principles, rather than simply being a heuristic that applies regardless of the situation.

Basic paradigm

In a series of experiments, we have taught or provided subjects with artificial categories and then asked them to draw inferences about a novel item drawn from one of those categories (Murphy & Ross, 1994, 2005; Verde, Murphy, & Ross, 2005). We typically have provided displays of the categories and some of their members, so that subjects do not have to rely on memory to make their judgments. (However, the results have been much the same whether we taught the categories to subjects before asking the questions or simply showed them the displays; Murphy & Ross, 1994, Experiment 3; Verde et al., 2005). We used this paradigm here, because, unlike other methods using natural categories (Ross & Murphy, 1996), it allows us to compare exemplar- and category-level information.

Consider Fig. 1, which shows the "drawings" of four children using a computer program that allows them to choose shapes and colors to make different figures. Here, the children are the categories, and their specific drawings are the instances. The children differ in the kinds of drawing they make, although there is also some variability. For example, Anna likes rectangles but has also drawn other shapes. The induction question involves a new figure about which we give partial information (e.g., its shape) and ask subjects to infer one of its other features (e.g., its color). Because people can see individual exemplars in the display, it is possible for them to use either general category properties (e.g., "Ann most often draws rectangles") or individual objects (e.g., "There are three black rectangles") in making inductions, as we shall demonstrate.

In the induction questions, subjects were told of one property of a *new* drawing. (The fact that the questions concerned different drawings than the ones displayed was emphasized, because if the drawing is one of the displayed ones, then use of conjunctions would be correct and use of category base-rates would be incorrect in answering the question.) Subjects were asked to say who most likely made the drawing (its category) and what other property it had (the induction). For example, suppose that you were asked about a new drawing in the shape of a heart. Most likely Karla drew it, as she drew five of the six hearts shown in Fig. 1. You would give your estimate of the likelihood that Karla drew this new drawing—perhaps five-sixths. Then you would be asked what color you thought this heart most likely was. In working out potential answers for this question, we will assume that you focus your attention on Karla's figures, as our past work has suggested (Murphy & Ross, 2007). However, none of the predictions requires this assumption. Looking at Karla's figures, you could count that she has made four blue figures, three orange ones, and one black figure. Therefore, you could choose blue as the most likely color, perhaps with a 50% likelihood (four of eight figures). This reasoning corresponds to the independence assumption: In making this judgment, you looked at Karla's figures in general—the category-level information—and did not focus on the given feature of heart, since shape and color are assumed to be independent. Once you have decided that Karla is most likely to have drawn this heart, the shape of the present figure plays no further role in the induction.

The conjunction strategy provides a different reasoning process. Given that the present figure is a heart, you might wonder what colors other hearts have. Four of Karla's hearts are blue, and one is orange, so the new figure is probably blue. We call this a conjunction strategy because the induction is based on the number of exemplars with the conjunction of the given and target feature (blue hearts). In this example, the answer you would get from assuming independence is the same as the one you would get by looking at the conjunction—blue in each case. This is what we call the *agree* condition, meaning that the same answer is obtained by the two strategies, base-rate and conjunction, because the most common color in the category happens to be the one that is also most commonly associated with the given feature.

In general, though, an answer made by focusing on the given feature need not be the same as the category-level answer. If the two features are independent, it will be, but if they are not, then the answer depends on the correlation between shape and color. This is illustrated by a different question. Imagine that you are told that a new figure is yellow. Maura has drawn five of the six yellow figures, so she is the most likely category. Furthermore, she has drawn four diamonds, three circles, and one heart, so the most likely shape is probably diamond, on the independence assumption. But if you used a conjunction strategy, you would find that, of the yellow figures, three are circles and only one is a diamond (plus a heart). Now the conjunction strategy gives a different answer from assuming independence, which yields the condition name *the disagree condition*: Going by category base-rates leads to the answer that the yellow figure is probably a diamond, but looking at the color-shape conjunctions leads to the inference that it is a circle.[1]

[1] The mild uncertainty about classifying the object was retained to maintain the similarity with our earlier work that had uncertain categorization. However, the item that was not in the target category could not much influence induction in either condition. For example, in the problem in which people had to predict the shape of a yellow figure, five of the six yellow figures were in Maura's category, and the critical question was whether people chose the shape circle or diamond. The sixth yellow item was drawn by Elif, who did not draw any circles or diamonds. So, the item in the other category would not have influenced the choice between these shapes whether people attended to it or not. As will be seen, subjects did not choose a shape from the unlikely category (Elif) in their inductions.

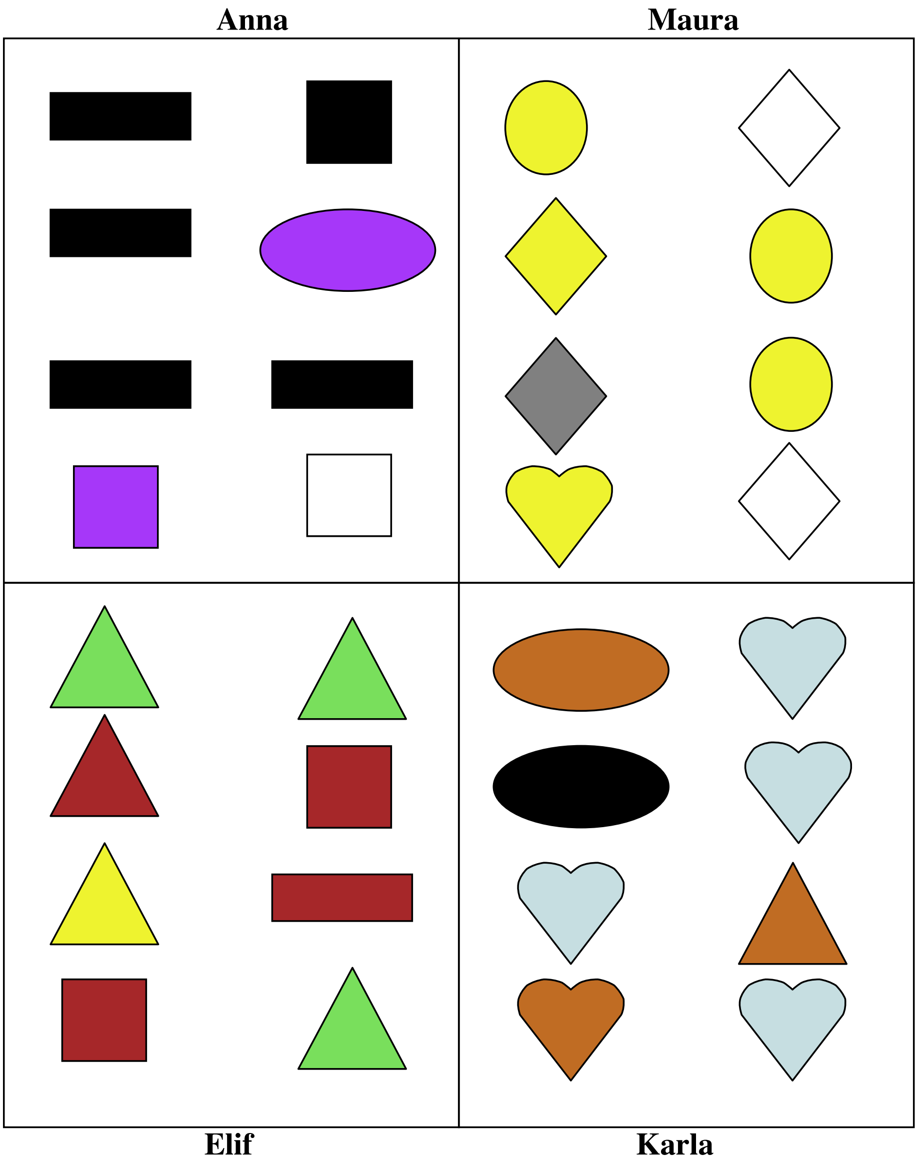

Fig. 1. An example of the materials used in Experiment 1, children's computer-assisted "drawings." The agree condition consisted of predicting the shape of a black figure and the color of a heart. The disagree condition consisted of predicting the color of a triangle and the shape of a yellow figure. These assignments were reversed in another version of the materials.

The logic of our experiments, then, is as follows. If people assume that features are independent, they will look at the base-rates of the relevant features in the target category and so will not be influenced by conjunctions. They will give the same answers in the agree and disagree condition: the most common feature in the category. However, if people restrict their examination to objects that have the given feature, they will give different answers in the two conditions: the most common feature in the agree condition, but the feature that was paired with the given feature most often in the disagree condition.

An important aspect of the design is that the agree and disagree conditions were identical in terms of within-category feature frequencies; the disagree condition was formed merely by re-pairing the features used in the agree condition. In Maura's category, the yellow figures are mostly circles. But in another version, the shapes were shifted around so that her yellow figures were mostly diamonds (to make an agree condition). If people treat features as independent, then shifting around the colors and shapes so that they are paired differently will have no effect, and people's inductions will be similar in the two conditions. If people make their inductions through feature conjunctions, they will give different answers in the two conditions.

Experiment 1

Experiment 1 carried out the design just described using children's drawings, as in much of our prior work. In later experiments, we expanded the stimuli to different, more realistic kinds of categories. Because pilot testing suggested that the results would be very clear, we took the strategy of using small numbers of subjects in each of the initial experiments, which explore the conditions under which feature independence is found. As will be seen, the results were reliable within each experiment and were also replicated across experiments.

Method

Subjects

Ten NYU undergraduates, randomly assigned to one of the two displays, received course credit for participating.

Materials

Two experimental displays of the sort shown in Figure ure1 were used. Each display contained four categories of children's drawings, purportedly made by different children. Each drawing was a colored shape. The two versions of the displays consisted of essentially the same figures, except that color and shape were re-paired to make the agree and disagree conditions as defined above. In the display shown, Maura and Elif served in the disagree condition, and Anna and Karla served in the agree condition. These assignments were reversed in the other version of the display.

A questionnaire was used to ask four induction questions. Each question had four parts, as shown in this example:

I have a new drawing that is yellow. What child is *most likely* to have drawn it?
What is the probability (0–100) that it was drawn by this child?
What shape do you think this yellow drawing has?
What is the probability (0–100) that it has this shape?

We did not have specific predictions concerning the final probability question. It was included primarily in case significant differences were not found in the induction judgment (question 3) itself. As we always found differences in the induction itself, we do not present the probabilities here. (In fact, the probabilities cannot be directly compared when people make different inductions across conditions, because the probabilities are of different answers.)

Procedure

Subjects read a general instruction sheet concerning the children's drawings and the displays. They were told that the drawings were a random sample from a much larger set of drawings the children made and that they would be asked questions about *new drawings*. We emphasized the newness of the drawings in both the instructions and questions so that subjects would not think that they were being asked about the displayed drawings themselves. The instructions also explained the 0–100% scale, saying that 0 meant that something was impossible, 50% meant that it would happen about half the time, and 100% meant that it is completely certain.

We constructed the questionnaires so that the first two questions were in the same condition and the last two in the other condition. As a result, half the subjects answered questions about agree conditions first and half about disagree conditions first. If there were order effects, this would allow us to use the answers from the first two questions, which would not be influenced by answers in a different condition. In the event, there were no such effects. Two questions used color as the given feature and asked about shape; two used shape as the given feature and queried color.

Results and discussion

There was only one trial in which an unexpected classification was made to the first question (e.g., when asked about a yellow figure, the subject gave an answer other than Maura). That trial was omitted from analysis. Subjects were generally confident in their classification, with a mean rating of 85. The critical data concern the induction made about the unknown feature. To allow comparison across conditions, we scored each answer for whether it was the most frequent property or not. According to the independence hypothesis, this is the correct answer, which should be given in both conditions. In fact, people chose the most frequent feature 95% of the time in the agree condition, but only 15% of the time in the disagree condition, $t(9) = 9.80$, $p < .001$). Indeed, every subject gave the most frequent answer more often in the agree condition than in the disagree condition, suggesting that every subject was using conjunctions.

Under an assumption of feature independence, our subjects were inconsistent: They seemed to be using the category base-rate in the agree condition but not in the disagree condition. However, if independence is not assumed, then their behavior is consistent: Subjects were looking at the target features only in items that had the given feature (e.g., looking only at shapes of yellow figures), which results in different answers in the two conditions. An examination of their responses in the disagree condition revealed that every time the category base-rate was not followed, the feature induced was always the one predicted by a conjunction strategy. A recent report by Papadopoulos, Hayes, and Newell (2009) also finds evidence for the use of conjunctions in induction, using a similar paradigm.

Experiment 2

Clearly, people are not treating the features as independent in making their predictions, though the rationale for it in the domain of children's drawings is mysterious. Is it really the case that children's drawings have strong correlations? Our own intuitions are that different children might like different shapes and different colors, but that they do not have strong feelings about the conjunctions. For example, a child might like green and red and also like triangles and circles. But would the child create many more green circles than red circles, and many more red triangles than green triangles? In the domain of children's geometric drawings, there does not seem to be a clear rationale for some colors to go with some shapes and not others. One could always think of a reason for such a correlation, but should one *assume* a correlation of shape and color in the population of children's drawings? And if there is no such correlation, then the category base-rate would likely be a more reliable source of information, as discussed in the Introduction.

However, it is conceivable that our intuitions about children's drawing are not shared by our subjects, so we decided to directly manipulate people's expectations about the relations among stimulus dimensions. To do so, the next three experiments used a different domain, where feature correlations seem more likely. In particular, the items were cars and their properties, and the categories were car manufacturers. Because these properties were often not easily depictable, we displayed the exemplars verbally, as pairs of features. Therefore, we first attempted to replicate Experiment 1 with new stimuli, to make sure that the feature conjunction strategy occurred with these new materials. These stimuli used stimulus dimensions that do not have any obvious relation to one another. In the subsequent experiments we introduced dimensions that do have relations to one another to see whether this influenced induction.

Method

Experiment 2 was very similar to Experiment 1, except that the categories were car manufacturers (Honda, Ford, Volvo, and BMW) and the items cars. Each car appeared as a pair of verbal features, one describing its color and the other the "free" feature that came with that particular car (e.g., satellite radio, alloy wheels, heated seat). The design of the stimuli was exactly as shown in Fig. 1, and the questionnaires also followed the same design as in Experiment 1. Thus, the experiment is a close replication of the previous experiment's procedure, but with very different stimuli. This change should also make clearer that the sample of exemplars in each category is really a sample, because people know that manufacturers make many cars. Ten subjects served in this experiment. They were randomly assigned to counterbalancing groups (and in fact were randomly assigned to the experiment, which was run in parallel with Experiments 3 and 4).

Results and discussion

Mean confidence in the initial classification was 82. Recall that the dependent measure is how often people choose the most frequent feature in a category. People chose this feature nearly all the time in the agree condition (95%) but seldom in the disagree condition (20%). This difference was reliable, $t(9) = 5.58$, $p < .001$. So, Experiment 1's results were replicated here.

Again, our intuition is the color and features are only very loosely related. Why should black cars have an iPod dock but purple cars have satellite radio? Although it is possible that there is such a connection, it does not seem to be common knowledge, so subjects should not assume that color and feature are in fact correlated. We addressed this issue systematically in Experiments 3 and 4.

Experiments 3 and 4

In Experiment 3, we used features that people would expect to be related in the domain of cars. So, rather than listing color, we listed another feature that was semantically related to the "free" feature. For example, one stimulus dimension of the Honda category was surround-sound system, 100-watt amplifier, or audiophile speakers. Each car had a value on that dimension. The other dimension was iPod dock or satellite radio. Because the two stimulus dimensions are conceptually related, people might expect there to be an association between features. For example, perhaps people with iPods want 100-watt amplifiers and people with satellite radios want surround sound, and therefore manufacturers pair these features when designing cars. Such dimensions should encourage use of the conjunction strategy. Note that we are not saying that people expect iPods to go with 100-watt amplifiers, but rather that people may expect there to be *some* correlation among these dimensions and therefore will use the conjunction strategy, which will be sensitive to whatever correlation there is.

In Experiment 4, we used color and feature of cars, with the exact same materials as in Experiment 2. However we developed a cover story in the main instructions that emphasized the independence of the two stimulus dimensions. The cover story explained that these cars were built in factories with robotic painting arms. The robot randomly selected the paint for a specific car, according to a schedule made by the manufacturer. For example, one manufacturer might decide that it wanted 25% white cars, 30% violet cars, 10% orange cars, etc., and the robot randomly chose the color for each car using these probabilities: "The robotic arms have no idea what the other features of the car are, of course. They only produce the color that has been randomly chosen for that car." These instructions made it clear that there was in fact no overall relation between color and any feature, within any given manufacturer's category. Also, referring to the painting device as an *arm* implied that it could not see the car's other features. Other aspects of the design were the same as in the earlier experiments. Ten subjects served in Experiment 3, and 11 served in Experiment 4, randomly assigned, as explained earlier.

Results, Experiment 3

This experiment emphasized the semantic relationship between the dimensions. Given that Experiment 2 has already shown that people have a strong tendency to use conjunctions, it is not surprising to find that they had the same tendency when the features were semantically related. (As Experiments 2–4 were run in parallel, the results of Experiment 2 were not known when Experiment 3 was performed.) Confidence in the initial categorization was 85. Subjects chose the predominant feature in the category 90% of the time in the agree condition but only 10% of the time in the disagree condition, $t(9) = 7.24$, $p < .001$. Nine of the ten subjects showed the expected effect (the other had no difference). Therefore, this is another replication of the use of conjunctions in induction.

Results, Experiment 4

This experiment emphasized that there was no relation between color and the other feature, because color was chosen randomly by a robot that did not know what other feature the car had. (And robots are well known not to care about surround-sound systems, etc.) Confidence in initial categorization was 69. Surprisingly, the results in this experiment were identical to those of the three previous ones. Subjects chose the most frequent feature 100% of the time in the agree condition but only 9% of the time in the disagree condition, $t(10) = 14.91$, $p < .001$. All subjects showed the effect. Indeed, although it is no doubt due to chance, the effect is numerically larger here than when the features were semantically related (Experiment 3). Clearly, then, people's use of the conjunction strategy does not seem to be driven by the belief that there is a correlation between the two features.

Experiment 4a

Given the surprising results of Experiment 4, we were concerned that subjects may not have fully understood what it meant to say that the robot chose the colors randomly. Furthermore, when they were doing the inductions, they may have forgotten the information about the robotic arm. Therefore, we took steps in a follow-up experiment to ensure that subjects understood this random process. Prior to taking the questionnaire, they read a description of the robot, as before. Then they were told that they would act like the robot themselves. They were given a pair of dice and told that they would throw the dice and then determine the color of cars, according to a written schedule of two categories not used in the main experiments (Mercedes and Toyota). Each manufacturer had six cars listed, along with its "free" feature. Subjects threw the dice for each car, looked at the colors listed for different dice values (e.g., for Mercedes, 1–6 was white, 7–10 was blue, and 11–12 was silver), and then wrote down the color that each car would be painted. The instructions explained that the robot used the identical system to determine color, only using a computer instead of dice to generate a random number.

This procedure ensured that subjects would understand that the colors for each car were determined randomly, and that the feature listed for each car had no relationship to the color whatsoever, given that they did not use the feature to determine the car's color. The procedure also made it very clear that the categories differed in their properties, as the schedules for the two manufacturers were different. Immediately following this practice, subjects completed the same four questions about new cars as in Experiment 4. It seems unlikely that people would forget this practice with the dice when answering the immediately following questions.

The results of this follow-up corresponded closely to those of the other experiments. Initial categorization was rated 73. Subjects chose the predominant feature 100% of the time in the agree condition but only 27% of the time in the disagree condition, $t(10) = 5.16$, $p < .01$. Eight subjects perfectly followed the conjunction strategy, and three always chose the most frequent feature regardless of conjunction. In short, extended, concrete experience with the random nature of the color assignment process did not prevent subjects from using conjunctions as if a car's color helped to predict its other features. It is possible that the three subjects who consistently used base-rates were influenced by the robotic dice training, but what is more remarkable is that eight of eleven subjects were totally uninfluenced by it. And these results suggest that the findings of Experiment 4 were not due to incomprehension of the randomness of the robotic arm.

Discussion of experiments 3–4a

After Experiment 1, we expressed doubt that there is a close relationship between the color and shape of children's drawings that would motivate a conjunction strategy. This strategy is most appropriate when one expects stimulus dimensions to be correlated. Especially in small samples, there would often be spurious correlations between dimensions that could strongly influence inductions when this strategy is used. Therefore, if no correlation is known to exist, simply using the category base-rate ("Most Fords are blue") would yield the most reliable answer.

Surprisingly, manipulating people's expectations and knowledge of feature relations had no effect whatsoever on their inductions. Telling subjects that one feature was selected randomly, without any knowledge of the other feature, did not prevent people from nonetheless using the conjunctions of those dimensions in making their inductions. Even giving concrete experience in assigning colors randomly had little effect. The results were identical when we emphasized the relations between dimensions, when we emphasized that there was no relation between dimensions, and when we did not specify any relation. Therefore, it does not seem as if people are using conjunctions out of a belief that dimensions are correlated, because they do so whether or not there is reason to think the dimensions are correlated. Instead, the conjunction strategy appears to be a heuristic that applies regardless of this statistical justification—and contrary to the widespread assumption that inductions are based on summary category-level information.

Experiment 5

The results so far support the idea that people use feature–feature relations in preference to category-level information in making inductions as a general heuristic rather than out of knowledge connecting the features or a general belief that they are correlated. We created a new questionnaire that would put this heuristic to a more severe test. In this induction question, we did not tell subjects a property of the specific item they were making an induction about. Instead, we described the property as being about the category as a whole. Perhaps people have experience in answering induction questions by looking for matches between the test object's features and known instances regardless of category membership. If you are trying to decide whether a furry growling thing bites, you might think of other furry growling things you have known—regardless of their categories—and consider whether they are biters (as Papadopoulos et al., 2009, propose). In contrast, using category-based induction and the assumption of feature independence, you would infer that the furry growling thing is most likely a cat, and then decide what proportion of cats bite. Perhaps the key to people's use of conjunctions is whether they truly consider the prediction to be based on a category or an individual.

In this experiment, we provided people with a feature of the *category* of the item rather than its own feature. That is, we told people that the car comes from a category (manufacturer) with a lot of yellow cars, rather then telling them that this car is yellow. If the conjunction strategy is based on the assumption that ''This car is yellow, so I need to look at other yellow cars to draw the inference,'' then that reasoning would be short-circuited in this problem: The present car's color is not known. Thus, if people are using feature conjunctions because they explicitly believe that object information is more important than category-level information, then this procedure should discourage them from using a conjunction strategy.

Method

Sixteen NYU students served in this experiment. The materials were the colored cars used in Experiment 2. However, the questionnaires differed in that the first question took the form: ''I have a new car that was made by a manufacturer who makes a lot of yellow cars. Which manufacturer do you think *most likely* made this car?'' Thus, the given feature was attributed to the category (manufacturer) rather than to the test object. Subsequent questions asked which other feature that car had, as before.

Results

The dependent measure, as before, is how often subjects chose the most frequent feature in the category. Previously, people did so in the agree condition but not the disagree condition, showing that they had used feature conjunctions. We expected not to find that pattern in this experiment. However, the pattern did again recur, albeit slightly less strongly. Subjects chose the most frequent feature 94% of the time in the agree condition and 28% of the time in the disagree condition, which was a reliable difference, $t(15) = 4.87$, $p < .001$. Confidence in initial categorization was 86.

If our expectation had been correct, we would have seen high proportions of choosing the most frequent feature in both conditions. Therefore, the rather low 28% figure in the disagree condition is the unexpected one. (All theories predict a high proportion in the agree condition, which was in fact at least 90% in every experiment.) That figure *is* higher than in any of the previous experiments, but it is not qualitatively different from the previous means, which were 15%, 20%, 10%, 9%, and 27% in this condition. Furthermore, 11 of the 16 subjects showed less frequent-feature responding in the disagree condition, suggesting that they were using the conjunction strategy. Four subjects seemed to be using the base-rates consistently (choosing the most frequent feature for all responses), and one seemed confused.

Discussion

The reworded question probably did have a small effect, as the difference between the two conditions was slightly less here than in prior experiments. Nonetheless, the most striking finding is that even in a case in which the conjunction strategy does not truly make sense, the majority of subjects continued to use it. The questions did not say anything about the features of the given object but only gave a clue as to which category was the correct one. Surprisingly, people latched onto that clue and treated it as if it were the feature of the given item. When told that the manufacturer made a lot of yellow cars, most subjects selected Honda as the correct manufacturer and then *looked only at its yellow cars* to decide the free option of the test vehicle, whose color was not given. It is rather surprising that subjects would consult the features of only yellow cars to infer the feature of a car that is not known to be yellow.

One might argue that people interpreted the question as indicating that the given vehicle was in fact yellow (or whatever), perhaps through some sort of Gricean inference. This is certainly possible, although the instructions stated that the feature was ''a clue about who the manufacturer might have been,'' rather than saying that it was a feature of the car. Indeed, our Gricean intuition is that if you say that the car was made by a manufacturer who makes a lot of yellow cars, you are specifically *avoiding* saying the car is yellow, and so the implicature would be that the car is not yellow. Furthermore, examination of the display would show that the manufacturer's cars are only half yellow, and so it is very possible that one of its cars has a different color. In short, there is no clear justification for people to infer that the feature described as true of the manufacturer is true of the particular car being asked about. If they drew this inference, it probably reflects their usual strategy in making inductions.

Experiment 6

One possible explanation of people's reliance on conjunctions is that it is an inference based on the evidence in the displays. In the agree conditions, the given and target features are in fact correlated. For example, in Anna's category of drawings, four of the five black figures are rectangles, and all of the rectangles are black. Of course, in a small sample with a few features, there are bound to be some random correlations just by chance. But perhaps subjects do not interpret these conjunctions as being due to chance. As one audience member said in a presentation of our first experiments, ''You're showing people all these correlated displays, and then you're surprised that they're paying attention to the correlation!''

We are skeptical of this proposal, because people use the conjunction strategy from the very first display, as shown by mean responses in the 90% range. For example, subjects in Experiment 1 chose the base-rate feature 80% of the time in the agree condition and only 20% of the time in the disagree condition on the very first question. It is hard to believe that people drew strong conclusions about the independence of stimulus dimensions from one category of eight items. However, our intuitions are not the strongest possible argument, and so we changed the displays in Experiment 6 to provide a simple test of this hypothesis. In particular, we constructed displays in which the stimulus dimensions of the first three questions were not strongly correlated. We then presented a disagree test item to see whether people would use the base-rate or the conjunction (which give different answers in the disagree condition).

We first determined how often conjunctions of the target and given features should occur by chance, in order to construct uncorrelated displays. The given feature occurred in five category exemplars (and one other item outside the category, which can be ignored). There were eight items in the category, and four of them included the target feature. How often, by chance, should these items also have the given feature; that is, how many conjunctions should there be? Basic probability theory reveals that two or three conjunctions are the most likely, each with probability one-third. One or four conjunctions are less likely, each with probability one-sixth. (Since there are only four occurrences of the target feature, there cannot be five conjunctions. With eight items, there also cannot be zero conjunctions.) The displays in the prior experiments have used the two less likely probabilities: The agree condition had four conjunctions, and the disagree condition had only one. Although not impossible, four such displays (out of four) would be statistically unlikely.

In Experiment 6, the first three questions asked about categories in which there were two or three conjunctions, which are most likely if the features were randomly associated. Therefore, by the time subjects got to the final question, they would have received evidence that test and given features were not generally correlated—they had the conjunctions that would be expected by chance. The final, test question was in the disagree condition. Table 1 gives an example of one of the uncorrelated categories (Anna) and the final test category (Karla). If the structure of our displays was responsible for our results, subjects should no longer use the conjunction strategy in the final question.

Table 1
Two categories from the materials of Experiment 6.

Anna		Karla	
Black rectangle	Black square	Blue oval	Orange heart
Purple rectangle	Black oval	Blue oval	Black heart
Black rectangle	Empty rectangle	Blue heart	Blue triangle
Purple square	Black square	Orange heart	Orange heart

Note. Anna's category represents the uncorrelated categories in which only two conjunctions between the critical features (black and rectangle) are present. Karla's category represents the usual disagree condition, in which heart is the given feature, and blue is the most common color (but more hearts are orange than are blue). The actual display used colored shapes and contained four categories.

Instead, they should choose the feature that is most common in the category.

Method

As described above, we made one new display of children's colored shapes for use in this experiment, similar to that shown in Fig. 1. In the first three questions, the critical category had two or three conjunctions between the given and target features. For example, there were two black rectangles (as shown in Table 1, Anna's drawings) and three red triangles (not shown). The final question used the given feature of heart shape and asked subjects to infer the color (see Table 1, Karla's drawings). There were four blue and three orange exemplars in the category, so use of the base-rate should lead to the answer "blue." However, there were three orange hearts and only one blue heart, so use of the conjunction should lead to the answer "orange." Twelve new subjects served in the experiment.

Results and discussion

The dependent measure was simply how often people gave conjunction responses to the final, test question (which had a mean categorization confidence of 77). If the alternative considered here is correct, then people will choose the base-rate feature. In fact, they chose the conjunction feature 82% of the time, and the base-rate item 18% of the time. (We omitted one subject who gave a nonstandard color name that did not refer to any color in the category and who also made an invalid response to an earlier question.) Thus, even after sufficient evidence that the features are independently distributed within the categories, subjects tended to use the conjunction strategy. Indeed, the effect is only slightly weaker than the original design (i.e., comparing 18% to the previous disagree condition means).

Another way to evaluate subjects' tendency to use the conjunction strategy is to compare their use of the base-rate feature in the first three problems with that in the final problem. The first questions were not identical to the previous agree condition, because they only had two or three conjunctions (instead of four). Nonetheless, if subjects were using conjunctions on those trials, we would expect them to often give the base-rate answer.[2] And, in fact, they did so 94% of the time in the first three conditions, which was reliably different from the 18% in the final, disagree question, $t(10) = 6.33$, $p < .001$, again giving evidence for the conjunction strategy. It is striking that using a manipulation that was quite weaker than the previous agree condition, we still found overwhelming evidence for the use of conjunctions in the first three questions.

[2] When there were three conjunctions, the conjunction strategy would yield the base-rate answer. When there were only two conjunctions, there were also two conjunctions of another feature. Subjects would have to use the category base-rate to break the tie in that case (see Table 1). Thus, the conjunction strategy alone could not always provide an unambiguous answer in these items, unlike the agree and disagree conditions used in prior experiments.

Thus, people's strong preference for a conjunction strategy cannot be explained by the statistical properties of our displays. Note that we did not set out to prove that display structure has no effect in general. It is possible, and even likely, that in a different paradigm subjects could learn the correlational structure of categories and use that to draw inferences (see Chin-Parker & Ross, 2002). Here we were testing the more limited hypothesis that in our previous experiments, people's use of conjunctions was a function of observing correlations in the displays. The present results, along with the observation of the conjunction strategy in the very first problem, rule out this explanation.

Experiment 7

People's use of the conjunction strategy seems remarkably resilient. Most surprising is the finding that the strategy does not appear to be based on a belief that the features are causally related or even correlated. For example, exposing people to displays where the features were not correlated (Experiment 6) did not eliminate the effect. Telling people that the features were created by different processes that were not connected (i.e., the robot painter) also did not eliminate the effect. Finally, we do not think that people have a strong belief that color and shape are correlated in children's drawings (Experiment 1) prior to doing the task. Using conjunctions seems to be a possibly unconscious strategy that people use to answer such questions in most circumstances, rather than a strategy that is explicitly selected because it is particularly appropriate for the materials or because of evidence that features are correlated.

How fixed is this strategy? Although a belief in the relatedness of the features does not appear to be necessary, does the use of the strategy depend upon a situation in which it at least makes sense to look for related features? An alternative is that this conjunction checking might be used because attention is directed to the first feature merely by its mention. Once people have a feature in their head ("yellow cars"), perhaps this feature then directs their information processing through unconscious priming: They begin to look at yellow items and count up their other features not because they believe that yellowness is in fact related to these features, but because the feature implicitly influences their attention and reasoning. "Yellow" is in their heads, and therefore they look at yellow things.

Experiment 7 tested this possibility in two ways, trying to find the limits of people's use of conjunctions. In the *category-name condition*, we no longer gave people the given feature name but instead gave them the category name. In all our experiments, people chose the target category at extremely high levels of accuracy, which is not surprising, given the structure of the displays. Therefore, providing the category was not giving people new information that they were not able to infer on their own. (Indeed, any trial in which a subject chose the wrong category was eliminated from analysis, since the objects in that category would not have the intended manipulation.) The first question no longer asked people to categorize the item, therefore, but to identify the *given* feature (which was no longer given to subjects, but we will retain this name to allow comparison with the earlier experiments).

An example will clarify. In previous experiments, subjects might be told, "I have a new car that is yellow. Who do you think manufactured it?... What feature do you think this yellow car has?" Thus, the given feature was explicitly provided to subjects, twice. They might answer that BMW manufactured the car and that it had mag wheels. In Experiment 7, the question would be, "I have a new car that was made by BMW. What color do you think this car is?... What feature do you think this car has?" People should be likely to provide the very common color yellow as the answer to the first question (and the data are analyzed only on trials when that answer is given). However, yellow is now their inference—it is not a known feature of the vehicle. Under such circumstances, it makes little sense for subjects to look only at yellow cars to identify the other feature, since they have not been told that the car is yellow. Instead, subjects may use the entire category of BMWs, which was the true given information.

In short, this condition will discover whether people will use a conjunction strategy even when the feature is not "given," but is their own inference. Why they might do so will be illuminated by the second condition.

The *multiple-item condition* was similar in that it also provided the category and asked subjects to infer both features. However, in this condition, the questions were split up so that they were about different items. For example, the corresponding questions to the above example would be, "1. I have a new car that was made by BMW. What color do you think this car is?... 2. I have another new car made by BMW. What feature do you think this car has?" The questions and their order were identical to those of the category-name condition, but the question numbering was changed so that the two parts of the original question (what category? what feature?) now became two separate questions asked about different cars.

Imagine that the reason that people use conjunctions is simply because the given feature value is directing their information processing in an unreflective ("automatic") way. That is, once they have the idea of a yellow car in their heads, they then only look at the features of yellow cars in order to make their inductions, even if there is no connection whatsoever between the color and the feature. If this behavior is truly unreflective, even mindless, then it should not make a difference that the color is not even of the same car. That is, the notion of yellow cars has been introduced by virtue of question 1, and so when you are asked about features of another car (in the same category) in question 2, you only look at the yellow cars. However, if the behavior reflects some more rational belief about induction, then the fact that the two cars are different should make people stop attending to the color. Once again, the difference between the agree and disagree conditions will be used as the measure of the use of conjunctions.

The category-name condition will test whether people use the conjunction strategy even if they provide the feature themselves, indicating a certain degree of irrationality in using this strategy, since there can be no argument that one should pay attention to conjunctions of features when one of the features is not truly known. The multiple-item condition will stretch this irrationality further, by seeing if the mentioning of a feature carries across to a brand new item. In previous work, we found evidence that people do not carry over information or strategies from one problem to the next (see Ross & Murphy, 1996), and so we predicted that there would not be any carry-over from one question to the next in the multiple-item condition. However, whether people will be affected by their own, inferred feature is not clear. Lagnado and Shanks (2003) found that people's guess about an item's category influenced their later inductions about it, even when they were given no information about the object to make the categorization. Perhaps, then, people's inference about the given feature will influence their judgment about the same item.

Method

Subjects

There were ten subjects in each group, plus two more eliminated for too many errors in choosing the "given" feature. (Since the issue of whether people will attend to conjunctions can only be answered when the correct given feature is selected, subjects who chose the wrong feature half the time or more were replaced.)

Stimuli

The overall displays and questions were closely based on the previous experiments using cars, their colors, and a "free" feature. Indeed, the categories and displays were the same as in Experiment 2. The differences were in the test questions. For the category-name condition, the questions took the following forms (exhortations to write legibly are omitted):

1. I have a new car that was made by Ford. What feature do you think this car has?What is the probability (0–100) that it has this feature?What color do you think this car is?What is the probability (0–100) that it is this color?

In half the questions, the free feature was the given property, and in half, the color was the given property. For the multiple-item condition, the corresponding question would be:

1. I have a new car that was made by Ford. What feature do you think this car has?What is the probability (0–100) that it has this feature?
2. I have another new car made by Ford. What color do you think this car is?What is the probability (0–100) that it is this color?

Thus, the only difference between the conditions was the change in question numbering and the reference in the second question to *another car*.

Procedure

The instructions and procedure were the same as in the previous studies.

Results

Although the experiment had two conditions, the issue was not a comparison of the conditions but rather whether people used the conjunction strategy in each one. Therefore, we analyzed each condition as in previous experiments, comparing inductions in the agree and disagree conditions. In the category-name condition, mean confidence in the predicted feature (first question) was 65. In the induction, the use of the conjunction strategy was greatly reduced from that of previous experiments, but it was not eliminated. People chose the feature with the highest base-rate 90% of the time in the agree condition and 65% of the time in the disagree condition. Although the latter statistic is much higher than in previous experiments (which ranged from approximately 9–25%), the two conditions were still significantly different, $t(9) = 2.24$, $p = .052$. Six of ten subjects showed the effect, with only one showing the opposite effect.

The multiple-item condition was identical, except that the question numbering and wording made the two questions (inference about the given and predicted features) appear to be about different objects. Clearly, there is no reason to think that when predicting a car's feature, one should look only at the color of the previously tested car. And, in fact, there was no difference between the two conditions in this case (*Ms* = 100% and 90% in the agree and disagree conditions). Here people did use the base-rates, even though their mean confidence in the category, 59, was no higher than in the category-name condition. Because almost all subjects chose the base-rate feature, it makes sense to compare their probability ratings of this feature. These probabilities did not differ across conditions either–the means were 50 and 48 for the agree and disagree conditions.

Discussion

The results from the category-name condition perhaps surprisingly show that about half of the subjects used a feature that they induced themselves, even though it was not certain, to restrict the information they used to answer a second induction question. Note that the most likely "given" feature (e.g., yellow in our Honda example) occurred in five of eight exemplars, and their ratings indicated that they were not certain that their best guess was correct. Even though subjects did not believe that their inference about the given feature was necessarily the right one, they often induced a less-frequent feature in the category, based on its conjunction with that feature.

This result is similar, then, to that of Lagnado and Shanks (2003), who found that asking people to guess an item's category, based on no given information, influenced a subsequent induction question. In both experiments, giving an answer about one question changed how people thought about the stimulus, and therefore how they answered subsequent questions about the item. The puzzle is that people knew that their first answer was not necessarily correct, and yet they acted as if it was. This finding is also quite similar to our own previous work showing that when people categorize an item, their subsequent inductions seem to assume that this category was known to be correct, even when they rate it as not being certain (Murphy & Ross, 1994, 2007, 2010; Ross & Murphy, 1996). However, we should note that only 60% of our subjects seemed to use the conjunction strategy under these conditions. Therefore, some people are not so unreflective in their use of the conjunction strategy but may have been sensitive to the fact that the given feature was inferred. On the other hand, only one subject consistently used the category base-rate, even though the category name was the information provided in the question.

One possibility is that people conditionalized their answer to the induction question based on their answer to the first question because of some kind of Gricean implicature. That is, if they guessed that the car was yellow, then they answered the question about what feature it had by assuming that the car was in fact yellow. We do not see anything in the questions that would encourage subjects to draw this implicature—instead, the fact that they have just rated the probability of their predictions as only 65% would seem to encourage them to not to assume the car was yellow. Furthermore, it is puzzling why answering a question about the feature should cause subjects to focus on the feature in this experiment, but answering a question about an item's category does not cause people to focus on category-level information in the previous experiments. Rather than saying that the questions somehow encouraged people to focus on the feature they inferred, we would say that the form of the questions did not *prevent* people from doing so if that was their strategy.

The multiple-item condition finally revealed some limits on people's use of the conjunction strategy. Their use of base-rates in the disagree condition here (90%) was comparable to results from the agree conditions. Simply having the feature "yellow" in one's head did not mean that one looks only at yellow cars. When the question about color concerned one car and the question about free feature another car, people did not use their inference about the first to inform their inference about the second, and so they gave an answer based on category-level information. This suggests that unconscious priming from mentioning a color does not explain people's propensity to use conjunctions. It seems then that people primarily use the conjunction strategy whenever they believe that the object they are making the induction about has a given feature—or is likely to have that feature. Merely having that feature activated from a previous question does not cause one to use it to draw inferences; the feature must be a feature of the object asked about.

General discussion

The present experiments show a strikingly strong preference for people to use a conjunction strategy when making predictions about the property of an object. These results are contrary to the assumption that features are treated as independent within categories, which we abstracted from Anderson (1991). However, the results also seem contrary to the general assumption of the entire literature on category-based induction, including our own past work, which has assumed that inductions are based on knowledge about a category's properties, such as ducks fly, or Hondas usually have good gas mileage. The present paradigm was designed to put category-level base-rates into competition with feature correlations, and therefore, any evidence for the conjunctions was necessarily evidence against category-level knowledge. When people chose the target feature that occurred most often with the given feature in the disagree condition, they were choosing a feature that was less frequent in the category. Therefore, the results do not merely show that conjunctions are one source of evidence that people use in situations like these, but that they in fact dominate the other statistical information available in these inductions.

One exception to the emphasis on category-level knowledge is Sloutsky and Fisher's (2004) theory of induction in children. They argue that induction from one object to another is a function of their overall similarity and shared category name. Thus, this model does put emphasis on exemplar-level relations in induction. However, it does not include category-level information at all, because the task only provides information about specific exemplars, not an entire category. Thus, the model cannot be applied to tasks in which entire categories are used in induction (e.g., Rips, 1975), and it does not explain how people might use category-level knowledge when it is available or information about specific objects is not known.

Much of our previous work (summarized in Murphy & Ross, 2007) explored how people make inductions about an object that has uncertain categorization (as in our initial growling animal example, which could be a dog, bear, or a number of things). The present experiments focused on items that are not very uncertain (e.g., a yellow figure is very likely drawn by Maura). This allows us to control the category-level information effectively. If we had used items with uncertain classification, then some people might use one category to make predictions, some might use the other, and some might try to use both, making it difficult to evaluate whether they were using category base-rates.

One consequence of using this design is that we cannot say what people might do if the items had an uncertain categorization. Conceivably, this could lead them to use category-level information rather than to rely on conjunctions. This seems unlikely, however: Why would people decline to use information about one fairly certain category but then use information about two uncertain ones? Papadopoulos et al. (2009) suggest that when items are uncertain, people still rely on feature conjunctions. One implication of these results is that in past work using similar paradigms, subjects could have been using a conjunction strategy, which might have clouded the results concerning whether they used multiple categories (but see Murphy & Ross, 2010).

Other approaches to category-based induction

We formed our predictions based on Anderson's (1991) independence assumption, but we also argued that much work on category-based induction seems to share this assumption if only by omission of any discussion of induction based on exemplar information or feature correlations within a category. The main exception is the work by Medin, Coley and colleagues, which suggesting that many inductions are based on specific knowledge of attributes and their relations, arguing that category-level induction may be a fall-back strategy used when specific knowledge is not available (Proffitt et al., 2000). Other approaches to category-based reasoning might have some flexibility to include feature correlations as a source of induction. One problem in applying such models to our task, however, is that they most often attempt to explain generalization of a novel feature from one category to another (e.g., If robins have X, do bats have X?), whereas our task requires inferring a known category feature to a novel object. Related to this, noticing and using feature correlations may require memory for individual exemplars, in order to use something like the conjunction strategy. We therefore focus on whether such models refer to exemplar-level information or to feature correlations in general.

Rogers and McClelland (2004) describe an ambitious project of representation of semantic knowledge, including a category-based induction component (chap. 7). Their model explicitly links features to categories and is sensitive to feature–feature relations through associations to shared categories. However, their model does not explicitly represent remembered exemplars: It has one node for canaries, one for daisy, and so on. Therefore, as it stands, it does not appear able to "peek" inside category-level knowledge (e.g., to see that male canaries have different colors than female ones) in the way our subjects did. However, with the addition of exemplar nodes between the features and category representations (e.g., along the lines of ALCOVE; Kruschke, 1992), the model could potentially be extended to account for our effects. Similarly, Sloman's (1993) feature-based model of induction could potentially account for feature correlations, although it would have to be modified from its current version, in which features are associated to concepts and not to one another. Currently, the model uses a measure of feature overlap of premise to conclusion categories and does not seem to use feature–feature relations.

Bayesian models of category-based induction have been proposed as well. So far as we know, none of them includes exemplar-level information in making inductions. Heit (2000, p. 590) pointed out that there was nothing in his model that required it to apply to categories rather than to individuals. However, his model does not use conjunctions or feature correlations, but rather more general similarity measures between categories.

More recently, Kemp and Tenenbaum (2009) described an impressive Bayesian approach to category learning and induction that addressed many different kinds of categories and forms of reasoning. One interesting aspect of their work is that they model very diverse forms of reasoning, such as causal reasoning and more traditional category-based induction. Therefore, it seems likely that their model has the power to do both category-level induction and induction based on feature correlations or other knowledge. However, further development may be required for them to simultaneously represent relations among features and category-level knowledge, as their reported examples all concern the latter (e.g., causal relations between predators and prey).

Any attempt to account for our results needs to be able to do two very different things: 1. the conjunction strategy, which requires exemplars to be readily available, and 2. "normal" category-based induction, when exemplar information is not available. So far as we know, no current model does both of these things simultaneously.

The real world

The significance of this result depends in part on whether features within natural categories are correlated. Recall Anderson's (1991) claim that the features of species are roughly independent within biological species and (perhaps) other categories. His reasoning seems to be based on the independence of genes in a "freely interbreeding species" population (p. 411). Although this may be true, there are in fact some well-known failures of feature independence in species that should be considered. The simplest way to document such correlations is to find distinctive subsets within a category. When a category has fairly well-defined subsets, items within each subset will share a group of attributes, and exemplars in the other subsets will differ in those attributes, yielding correlated attributes.

One example of such subsets is sexual dimorphism. Differences between sexes can be enormous, perhaps the most disgusting being that of the male osedax ("bone-eating" or "zombie" worm), dozens of which actually live inside a female conspecific, where they never develop beyond the larva stage (Rouse, Goffredi, & Vrijenhoek, 2004). More visible examples might include deer, wild fowl, spiders, many passerine bird species, and some fish. Male deer are larger than females and have antlers. The two differ in various behavioral traits, such as the fighting for mates and protective behaviors. Thus, size, antlers, and coloration differences are correlated in many deer species, and these are all correlated with behaviors and sexual organs. The more different the sexes of a species are, the more there will be intraspecies correlations of attributes, because the female properties will cluster closely together, separate from the male properties. Therefore, if you know that a deer had antlers, you might predict different properties of it than if you knew it did not.

Another example of feature correlations in species arises from developmental differences. Juveniles sometimes have quite different properties than those of adults, and therefore they create a set of correlated properties. Size and behavioral properties are often different for juveniles, which also can have different coloration (e.g., a fawn's spots or the dull feathers of a cygnet). Plants also may have such correlations, as when tree shoots are weak, lack branches, and have smooth bark, in contrast to large, strong, rippled old trees.

The most radical dimorphisms are found in animals that metamorphose, such as caterpillars or flies. Here, size, shape, body parts, behaviors, and habitats are all drastically different between the two life forms. These result in clusters of properties in each life form that are different in the other, causing strong feature correlations within the species (Murphy & Rosengren, 2010).

In short, the independence of features in species is not a general property that one can count on. Within subsets of species, such as adult male white-tailed deer or swamp maple seedlings, there may be greater independence of features. However, that may simply be a factor of homogeneity: When there is little feature variation, there cannot be strong feature correlations. One might suggest that the problem of feature correlation does not arise in real life, because people can simply operate at the level of more specific categories, like adult male white-tailed deer, where correlations are not a problem. However, the level at which people usually classify objects typically is much higher than these specific categories (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Indeed, Coley et al. (1997) found that American college students prefer to make inductions at the level of the genus (e.g., *oak, salmon, lily*) rather than at the level of the species (which students are often unfamiliar with). Thus, it is unrealistic to argue that feature correlations in biological categories are not relevant because features are not correlated at highly specific categories, when many people do not use or even know these categories.

Although we have focused on species here, people clearly deal with many other kinds of categories in their lives. How does the feature independence idea hold up for other categories, such as those of people, events, and artifacts? People and event categories seem to have the same problems with feature independence seen in plants and animals—at most levels there are subcategories that lead to feature correlations. The features of event categories such as football games depend much on whether they are at the high school, college, or professional level. Within artifacts, our intuition is that feature correlations are the rule. For example, looking at computers in our homes and laboratories, we find that functionality, screen size, software, core memory, and hard disk size are all correlated, mostly due to continuing improvements in all these dimensions, as well as correlations induced by price differences (the more expensive units having better values on most dimensions). Even a fairly homogeneous category such as books show relations between the content of the books, their size, types of covers, cost, and publishers. Social categories are extremely various, but there are surely some that have strong feature correlations, such as professions or ethnic groups. Professors can be subcategorized in various ways, such as by their discipline, type of research activity, type of institution, age, salary, and gender, with some clear violations of feature independence.

We have been focusing here on whether features are in fact correlated across exemplars—that is, actual objects—in a category. A related but not identical question is whether people believe that there are such correlations in categories. Malt and Smith (1984) obtained feature listings for categories within six superordinate categories. They discovered a large number of significant feature correlations within those superordinates (e.g., 54 for birds; 14 for flowers). For example, clothing worn in bad weather tends to be made of wool, and birds that sing tend to be small and eat insects. They also identified negative correlations, such as furniture with four legs tending not to appear on walls. Interestingly, although the correlations were identified from feature listings, when people were confronted with these correlated feature pairs, they judged only a third of them to be significantly related. When investigating the conjunction strategy in natural categories, it will be interesting to compare the features that are explicitly identified as correlated with those that are in fact correlated but not judged as related (as suggested by Ahn, Marsh, Luhmann, & Lee, 2002).

In short, we would argue that the world is full of correlations. Categories developed in part to take advantage of such correlations, by dividing things up into the coherent clusters of features (Anderson, 1991; Rosch, 1977). But it nonetheless seems to be the case that some correlations also hold within categories.

Combined with our results, this could lead to a proposal that is opposite to that of the independence assumption, namely that given that there are so many correlations, people should be expected to use feature–feature relations when making inductions. If small dogs are noisier than big dogs, then why shouldn't one use the size of a dog to help predict whether it will bark? This normative argument for the use of correlations is inconsistent with our result that people use them even when the population features are not correlated (Experiments 4, 4a, and 6) and when the object does not necessarily have the given feature (Experiment 7). Therefore, if one wishes to use environmental structure to motivate people's use of this strategy, it must be similar to the motivation for many heuristics, namely that it often improves decisions, but that people are not sensitive to when it is not appropriate and so overuse it.

Choice of strategies

One motivation we had for doing this research was to discover people's beliefs about feature relationships in real categories. Do people believe features are independent? By examining when people did and did not use the conjunction strategy, we expected to gain insight into whether and when people believe in feature correlations. Ironically, our results were so strong that they did *not* shed light on this question, because use of conjunctions was independent of people's beliefs.

This leads to our next main conclusion, that use of the conjunctive strategy is not consciously selected in order to achieve goals, that its use is largely (but not entirely) unreflective. People use the conjunction strategy whether or not they believe the features are correlated (Experiments 3–4a), whether or not the features are actually correlated in experience (Experiment 6), and for some subjects, whether or not the feature is known to be a feature of the target object (Experiment 7). It is doubtful that they are consciously thinking, "There's no relationship between the car's color and feature, so I'll only look at the green ones when trying to decide which feature this car has," or "The most popular Honda color is yellow, so I'll only look at yellow cars even though I don't know this car's color." It seems more likely that the use of the given feature to limit the items sampled reflects an entrenched heuristic rather than a conscious choice, which would make little sense in some conditions. Informal questioning of demonstration subjects (in lab meetings and talks) often revealed an inability to explain why conjunctions were thought to be important. When it was pointed out that a participant did not choose the category's most likely feature, there was a strong tendency for people to repeat, with some emotion, "But you *said* it was a triangle," as if this entailed that only other triangles should be considered. Further questioning revealed that most people never considered using the category base-rate instead of looking at conjunctions. It is not that people chose the conjunction strategy, using the object's features, over category-level information, but that category-level information simply did not seem to be relevant once an object's feature was provided.

This finding is somewhat surprising given that the first question of the standard induction problem is to identify the category, for example:

> I have a new drawing of a triangle. What child is *most likely* to have drawn it? [Category question]
> What is the probability (0–100) that it was drawn by this child?

In the past, we have heard concerns from readers and audience members that such questions bias subjects to pay attention to the category they write to answer this question and to ignore other categories. It has been suggested that there is an implication in this question that once the subject writes down "Elif," then he or she is committed to assuming that Elif is in fact the correct category and to act upon this basis. There is evidence that the initial question may have this effect on some subjects (Hayes & Newell, 2009; Murphy & Ross, 2010).

In light of these concerns, it is surprising that when we use the same question in this experiment, and people write down the same category, they do not use category-level information for their inference. That is, if there is a commitment to the category as a result of writing down the category name, there should be a bias to increase use of that category, and yet we found a strong preference not to use category-level features. People may use the category in a different way, however, namely to restrict the items they apply the conjunction strategy to. That is, the category name does not seem to make people use category-level information, but it might have made them look only at conjunctions in the target category (though a recent experiment by Papadopoulos et al. (2009) suggests that even this use of categories is uncommon).

One reason that people may have for using the conjunction strategy is the belief that specific information is always more predictive than general information. Once you know a property of an object, that is a concrete fact about that particular object, which therefore may seem more relevant than information about an abstract category. In many cases, this assumption may be true. If you know the age of the oak in your back yard, you may be able to predict its strength more accurately than simply knowing the strength of oaks in general. However, this assumption is counterproductive when the given, concrete information does not actually provide specific information about the other feature, above and beyond identifying the category. In such cases, it may reduce accuracy, because category-level information is more reliable.

Other situations and stimuli

Our experimental paradigm used artificial categories that were continuously present, raising the question: In what other situations would people use this strategy? Obviously, using feature–feature relations to make inductions requires one to know or to be able to calculate those relations. In our displays, people could simply look at the items and count up conjunctions of the given and target features. Thus, information about feature relations was readily available. For familiar semantic categories, this strategy might not always be possible, as one is less likely to have detailed exemplar representations of squirrels, Camrys, or plumbers to allow one to make such predictions. More likely, the necessary information will be part of general world knowledge, such as tree experts' knowledge of disease transmission (Proffitt et al., 2000).

Although there are major difficulties in adapting our design to natural categories, due to differences in individuals' knowledge about potentially correlated features (and the need to find inductions where the feature-based answer is different from the category-level answer), it would be very interesting to investigate whether feature correlations are used in natural categories. It is possible that category-level and feature–feature inferences compete in inductions where they are both available. For example, imagine that you see an angrily barking labrador retriever. On the one hand, the dog's features suggest that it is aggressive and likely to bite; on the other, labs are well known to be friendly and mild. In such a case, both sources of information may sum to yield a compromise inference, such as crossing the road but not actually running away from the dog.

It is also important to consider that in some situations inductions must be done quickly, and so efficiency may determine which mechanism is used. In our paradigm, subjects had as much time as they wished to count up exemplars with the given feature and make up their minds. In other situations, in which exemplars must be retrieved from memory and various features retained in working memory, the conjunction strategy may not be fast enough to provide a reliable solution in the time available. Perhaps then people will tend to rely on category-level knowledge. Research on decision making shows that people's strategies do change with time pressure, using more complete comparisons when time is plentiful (Payne, Bettman, & Johnson, 1988). Time pressure influences category-based induction as well (Verde et al., 2005). In our example of the growling animal outside the tent, people might feel an urgency to make an immediate decision about whether to run away or stay inside, and so they might make a quick guess about its category and then decide accordingly (e.g., bear—run, raccoon—stay and fight), rather than engaging in exemplar retrieval and calculation of conjunctions. Alternatively, perhaps people simply access feature associations (growling—run) as the basis for their induction. Research will have to determine how time pressure changes inductive strategies.

Conclusion

In many situations, people classify objects based on their available properties and then want to make a further induction, such as whether a particular Honda is safe or that raccoon is dangerous or this lawyer is smart. Our results suggest that people may use correlations between the observable properties of the new object and the target feature, rather than relying on category base-rates to make the induction. When there are strong correlations between such properties within categories, then such a strategy is reasonable. But people use the strategy whether or not they believe such correlations exist, suggesting that induction is not optimal.

Although Anderson (1991) has been the main theorist to argue explicitly for feature independence, the implications of our results go beyond his particular proposal to theories of category-based induction as a whole. The assumption has been widespread in the field that category-level information is the basis of such inductions. Our results suggest that feature–feature relationships are at least as important in inductions about individual objects, perhaps because people prefer to use concrete, specific information to draw inductions over abstract, category-level information, even when the latter is more reliable than the former.

Acknowledgments

This work was supported by NIMH Grant MH41704. We thank Erin Jones, Bob Rehder, and Eric Taylor for helpful comments on an earlier version of this article.

References

Ahn, W.-K., Marsh, J. K., Luhmann, C. C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition, 30*, 107–118.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409–429.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition, 30*, 353–362.

Coley, J. D., Medin, D. L., & Atran, S. (1997). Does rank have its privilege? Inductive inferences within folkbiological taxonomies. *Cognition, 64*, 73–112.

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

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development, 58*, 1532–1541.

Hayes, B. K., Ruthven, C., & Newell, B. R. (2007). Inferring properties when categorization is uncertain: A feature-conjunction account. In D. S. McNamara, & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 1073–1078). Austin, TX: Cognitive Science Society.

Hayes, B. K., & Newell, B. R. (2009). Induction with uncertain categories: When do people consider the category alternatives? *Memory & Cognition, 37*, 730–743.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7*, 569–592.

Heit, E., & Barsalou, L. (1996). The instantiation principle in natural categories. *Memory, 4*, 413–451.

Kaplan, A. S., & Murphy, G. L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 829–846.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116*, 20–58.

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

Lagnado, D. A., & Shanks, D. R. (2003). The influence of hierarchy on probability judgments. *Cognition, 89*, 157–178.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior, 23*, 250–269.

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

Murphy, G. L., & Rosengren, K. (2010). The two-body problem: Classification and reasoning about polymorphs. *Cognitive Studies, 17*, 6–35.

Murphy, G. L., & Ross, B. H. (2010). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 263–276.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27*, 148–193.

Murphy, G. L., & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition, 95*, 175–200.

Murphy, G. L., & Ross, B. H. (2007). Use of single or multiple categories in category-based induction. In A. Feeney & E. Heit (Eds.), *Inductive reasoning: Experimental developmental, and computational approaches* (pp. 205–225). Cambridge: Cambridge University Press.

Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science: Vol. 2. Theory and applications* (pp. 23–45). Chichester: Ellis Horwood.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.

Papadopoulos, C., Hayes, B. K., & Newell, B. R. (2009). Non-categorical approaches to property induction with uncertain categories. In N. A. Taatgen, & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society*. Austin: Cognitive Science Society.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534–552.

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 811–828.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14*, 665–681.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.). *Advances in cross-cultural psychology* (Vol. 1, pp. 177–206). London: Academic Press.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 736–753.

Rouse, G. W., Goffredi, S. K., & Vrijenhoek, R. C. (2004). Osedax: Bone-eating marine worms with dwarf males. *Science, 305*, 668–671.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166–188.

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

Verde, M. F., Murphy, G. L., & Ross, B. H. (2005). Influence of multiple categories in inductive inference. *Memory & Cognition, 33*, 479–487.