

# Journal of Experimental Psychology: Learning, Memory, and Cognition

## Feature-Based Versus Category-Based Induction With Uncertain Categories

Oren Griffiths, Brett K. Hayes, and Ben R. Newell

Online First Publication, November 7, 2011. doi: 10.1037/a0026038

### CITATION

Griffiths, O., Hayes, B. K., & Newell, B. R. (2011, November 7). Feature-Based Versus Category-Based Induction With Uncertain Categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi: 10.1037/a0026038

# Feature-Based Versus Category-Based Induction With Uncertain Categories

Oren Griffiths, Brett K. Hayes, and Ben R. Newell  
University of New South Wales

Previous research has suggested that when feature inferences have to be made about an instance whose category membership is uncertain, feature-based inductive reasoning is used to the exclusion of category-based induction. These results contrast with the observation that people can and do use category-based induction when category membership is known. The present experiments examined the conditions that drive feature-based and category-based strategies in induction under category uncertainty. Specifically, 2 experiments investigated whether reliance on feature-based inductive strategies is a product of the lack of coherence in the categories used in previous research or is due to the use of a decision-only induction procedure. Experiment 1 found that feature-based reasoning remained the preferred strategy even when categories with relatively high internal coherence were used. Experiment 2 found a shift toward category-based reasoning when participants were trained to classify category members prior to feature induction. Together, these results suggest that an appropriate conceptual representation must be formed through experience with a category before it is likely to be used as a basis for feature induction.

*Keywords:* induction, category learning, concept formation, category coherence

Inductive inference is a core competency of intelligent agents. It allows people to leverage prior knowledge and experience to make predictions about unobserved instances or events. Much previous research has focused on how category knowledge is used to make predictions about the features of exemplars when their category membership is known (Hayes, Heit, & Swendsen, 2010; Sloman & Lagnado, 2005). This ability to infer hidden or novel properties of exemplars has been repeatedly demonstrated in studies using both real world (e.g., Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975) and artificial categories (e.g., Rehder, 2006; Yamauchi & Markman, 2000). For example, knowing that a person is a registered Republican may allow confident predictions of their views on social issues, such as gay marriage. What is less well understood is how feature induction works when an object's category membership is uncertain (e.g., if you were unsure about your friend's party affiliation). If you were only 70% confident that someone was a Republican, how could you leverage this (uncertain) information to inform your predictions about that person?

Initial work on this topic (Murphy & Ross, 1994) suggested that, just like the case in which category membership is certain, feature induction judgments under category uncertainty are based on categorical knowledge. However, recent studies using more sophisticated controls have shown that, when category membership is uncertain, feature induction judgments are predominantly based on a different source of information: feature conjunctions (Griffiths, Hayes, Newell, & Papadopoulos, in press; Murphy & Ross, 2010a; Papadopoulos, Hayes, & Newell, 2011). To illustrate, suppose you

knew that your colleague supported a woman's right to abortion but opposed government-subsidized health care. Although it is not clear which party your friend is affiliated with (the first view seems more consistent with liberal political views, the second with conservative views) you might predict that they would also have a positive view of gay marriage. This method of inference is referred to as a *feature conjunction* because it is based on the assumption that certain features (e.g., positive attitudes toward social issues like gay marriage and abortion rights) tend to co-occur.

This raises the important question of what drives the observed shift toward feature-based reasoning (and away from category-based reasoning) in induction with uncertain categories? One possibility is that the mere introduction of uncertainty leads people to see category membership as an unreliable basis for prediction; hence, they shift to feature-based inference. Alternatively, it may be that people's preference for different inductive strategies in tasks with exemplars whose category memberships are uncertain, as compared with tasks in which category membership is known, reflect differences in methods used to study these phenomena. An examination of studies of induction with category uncertainty (e.g., Murphy & Ross, 1994, 2010a, 2010b; Papadopoulos et al., 2011) suggests at least three major factors that might work against category-based reasoning: the use of weakly structured categories that lack internal coherence, the absence of experience with the stimulus categories prior to induction, and the ready availability of exemplar information during the induction process. The present experiments manipulated category coherence and the use of a decision-only methodology (which addresses the remaining two factors) to assess the impact of each of these factors on feature-based and category-based induction.

---

Oren Griffiths, Brett K. Hayes, and Ben R. Newell, School of Psychology, University of New South Wales, Sydney, Australia.

Correspondence concerning this article should be addressed to Oren Griffiths, School of Psychology, University of New South Wales, Sydney, Australia (2052). E-mail: oren.griffiths@unsw.edu.au

## Feature-Conjunction Strategies

Several recent studies of inductive reasoning have found that people frequently use feature-conjunction reasoning when they are

uncertain about an object's category membership (Griffiths et al., in press; Hayes, Ruthven, & Newell, 2007; Murphy & Ross, 2010a; Papadopoulos et al., 2011). These studies have used variants of the decision-only paradigm originally developed by Murphy and Ross (1994). In this procedure, participants are shown a feature of a probe exemplar that could belong to more than one category and are asked (a) to identify which category it is most likely to come from and (b) to predict its value on an unobserved feature dimension. To assist with these predictions, participants are shown a sample of the exemplars from categories to which the probe could belong. Because the primary interest in these studies was in inductive decision processes rather than category learning, category exemplars remained in view while inductive predictions were made.

An example (using the present stimulus set) is shown in Figure 1A, where the alternative categories are presented as species of bugs that live in different regions. In the example, participants are shown that a novel bug (the probe) has wings of a particular shape (hereinafter *downswept wings*; the given feature). They are then asked to infer which type of mandible that bug possesses. Note that the category membership of the probe is uncertain, as two different categories contain bugs with wings of the given shape.

When faced with tasks of this kind (hereinafter termed *uncertain induction* tasks), people overwhelmingly use feature-conjunction reasoning (Papadopoulos et al., 2011). Figure 1B outlines the details of the inductive predictions on the basis of feature-conjunction and category-based reasoning for the example shown in Figure 1A. As shown in Figure 1B, feature-conjunction reasoning could be instantiated in two ways. First, participants could identify the most likely category to which the probe bug exemplar could belong and then could examine only the bugs from that category (a *single-category feature-conjunction* strategy). When applied to the Figure 1A example, this strategy leads to a focus on bugs from the western region (which had the highest proportion of downswept wings) and a prediction of pointy, two-pronged mandibles (the first response option listed in Figure 1B). Alternatively, participants could ignore the category structure entirely and examine all bugs with downswept wings to determine the most likely type of mandible irrespective of category bounds (the *multiple-category feature-conjunction* strategy). As shown in Figure 1B, this strategy leads to the prediction of four pronged mandibles (the second response option in Figure 1B).

Papadopoulos et al. (2011) and Griffiths et al. (in press) found a strong preference for the multiple-category feature-conjunction strategy (although Griffiths et al., in press, also observed a minority that consistently used the single-category feature-conjunction strategy). For example, Papadopoulos et al. (Experiment 1) found that 97% of feature predictions were consistent with multiple-category feature-conjunction (also see Murphy & Ross, 2010a).

### Category-Based Strategies

These feature-based strategies can be contrasted with reasoning strategies based on category-level information. For example, Murphy and Ross (1994, 2007) have suggested that when faced with the task of feature induction under category uncertainty, people may solve the induction problem by using category-level information and ignoring the uncertainty associated with the categorization process. Turning to our original example, if there was a 70%

chance that your friend was a Republican, then you might reason as if that classification were certain, resulting in a prediction of conservative views on marriage. In the analogous example shown in Figure 1A, this single-category inductive strategy would first identify that bugs with down-facing wings were most likely from the western category (hereinafter the *target category*). Then, the most frequent type of mandibles of all bugs in the western category (including those with other types of wings) would be tallied. This would lead to a prediction of blunt mandibles (the third response option in Figure 1B).

There is, in principle, a fourth option. Bayesian approaches to category-based inductive prediction (e.g., Anderson, 1991) suggest that the optimal strategy is to consider the conditional probability of the property (e.g., supports same-sex marriage) in both more and less likely category alternatives (e.g., both Democrat and Republican classifications). Under this multiple-category strategy, these conditional probabilities are weighted according to the likelihood of each possible category assignment and then combined to yield a final prediction. In the Figure 1A example, this process would involve calculating the probability associated with each type of mandible within each category and then multiplying these values by the likelihood that a bug with downswept wings was drawn from each category. The calculations for this strategy are not shown in Figure 1B because this strategy was not explicitly examined in the present experiments (although it is considered in the General Discussion). This is because, to the best of our knowledge, there is no clear evidence that people use this multiple-category strategy.

Indeed, there is little evidence for category-based induction (single- or multiple-category variants) in uncertain induction tasks when feature conjunction is a viable alternative (Griffiths et al., in press; Murphy & Ross, 2010a; Newell, Paton, Hayes, & Griffiths, 2010; Papadopoulos et al., 2011). The sole example of clear use of a single-category strategy comes from the final experiment of Murphy and Ross (2010a). In this study, two probe exemplars were presented on each trial. For the first, participants were given a probe feature and were asked to identify its target category. For the second, participants were given the probe's category and asked to predict an unobserved feature. For this second probe, however, no feature information was given, making use of feature conjunction extremely unlikely. Only in the latter case did Murphy and Ross observe a majority of participants using single-category reasoning over the single-category feature conjunction. This finding clearly shows that people can use a single-category inductive strategy to solve an uncertain induction task when feature-conjunction-based strategies are rendered implausible. This result does not, however, explain why feature-conjunction strategies are typically preferred over category-based strategies under category uncertainty.

As noted earlier, it is possible that the mere introduction of category uncertainty encourages people to favor feature-conjunction strategies over category-based induction. Alternatively, the dominance of feature-conjunction strategies could be a product of the methodologies used to examine feature induction under category uncertainty. We now discuss three specific aspects of this methodology that could promote feature-conjunction reasoning: the use of relatively low-coherence categories, the absence of experience with these categories prior to inductive judgment, and the availability of exemplars at test.

**A**

Western Region                      Central Region                      Eastern Region

Suppose you found a sample of bugs and that most of these bugs had this type of wings:

First select which antennae/mandibles these bugs are most likely to have. Then rate how likely it is that these news bugs would have each type of antennae/mandibles.

**B**

Given the bug has this type of wings  ,			
which kind of mandibles do you think this bug has?			
Response Options	Single Category	Single-Category Feature Conjunction	Multiple-Category Feature Conjunction
	$= p(\text{antennae type 1}   \text{Western})$ $= 5/16$ $= 0.31$	$= p(\text{antennae type 1}   \text{Western} \& \text{wings})$ $= 5/10$ $= 0.5$	$= p(\text{antennae type 1}   \text{wings})$ $= 5/15$ $= 0.33$
	$= p(\text{antennae type 2}   \text{Western})$ $= 2/16$ $= 0.13$	$= p(\text{antennae type 2}   \text{Western} \& \text{wings})$ $= 2/10$ $= 0.2$	$= p(\text{antennae type 2}   \text{wings})$ $= 7/15$ $= 0.47$
	<b><math>= p(\text{antennae type 3}   \text{Western})</math></b> <b><math>= 9/16</math></b> <b><math>= 0.56</math></b>	$= p(\text{antennae type 3}   \text{Western} \& \text{wings})$ $= 3/10$ $= 0.3$	$= p(\text{antennae type 3}   \text{wings})$ $= 3/15$ $= 0.2$
	$= p(\text{antennae type 4}   \text{Western})$ $= 0/16$ $= 0$	$= p(\text{antennae type 4}   \text{Western} \& \text{wings})$ $= 0/10$ $= 0$	$= p(\text{antennae type 4}   \text{wings})$ $= 0/15$ $= 0$
	$= p(\text{antennae type 5}   \text{Western})$ $= 0/16$ $= 0$	$= p(\text{antennae type 5}   \text{Western} \& \text{wings})$ $= 0/10$ $= 0.33$	$= p(\text{antennae type 5}   \text{wings})$ $= 0/15$ $= 0$

Figure 1. Panel A shows an example of a high-coherence uncertain induction trial used in Experiment 1. All 48 exemplars (divided into three categories) were shown on-screen. An incomplete probe exemplar (with one given feature) was shown at the bottom of the screen. Participants were required to predict a missing feature of the probe exemplar. Note that bugs with the given feature (shown at the bottom of the screen) were most common in one category (western region), were less common in a second category (central region), and never occurred in the third category (eastern region). Panel B shows example calculations of probabilities of each possible feature prediction, using each of the three induction strategies. This example uses the category structure shown in Figure 1A (but also applies to that shown in Figure 2). The participants are told that they have a bug with a particular type of wings and are asked to predict the most likely type of mandibles. The most likely feature predictions based on each reasoning strategy are shown in bold.

### Category Coherence

Category coherence refers to the extent to which category exemplars and/or features are seen to go together in light of theoretical or causal knowledge (Murphy & Medin, 1985). Members of coherent categories typically show high levels of similarity to one another (Haslam, Rothschild, & Ernst, 2000), and people are more likely to generalize a property from a high-coherence category than from a low-coherence category to a new member (e.g., Hayes, Kurniawan, & Newell, 2011; Patalano & Ross, 2007; Rehder & Hastie, 2004).

Many studies of induction with certain categories (e.g., Osherson et al., 1990) have used natural categories that are well documented as being at least somewhat coherent (Hampton, 1998; Sloman, Love, & Ahn, 1998). Likewise, the few studies that have used artificial stimuli to examine induction (e.g., Yamauchi & Markman, 2000) have generally used family resemblance categories where most category members share a majority of features in common.

In contrast, an informal examination of the categories used in previous studies of uncertain induction indicates that they typically lack coherence (see Experiment 1 and Appendix A for more detail). More concretely, several previous studies of uncertain induction (e.g., Murphy & Ross, 1994; Papadopoulos et al., 2011) have used categories whose exemplars are abstract geometric forms. As a general rule, the exemplars used in these categories do not tend to share common features (such as color and shape) with the other exemplars from the same category, and conversely, exemplars of one category are not clearly distinct from exemplars drawn from alternative categories. Because there is little to unify the exemplars of each category, the category labels provide relatively little information about individual exemplars. To some extent, this is a by-product of the constraints imposed by the research question. To study induction with category uncertainty, there must necessarily be some overlap between the features of the members of at least two categories, and ideally, a particular feature value (the given feature, e.g., downswapped wings) must be associated with a number of other features (e.g., types of antennae) to generate different responses for each of the possible induction strategies. These design features reduce coherence.

Nevertheless, this lack of coherence may itself reduce the likelihood that people will use category-based induction. More specifically, if knowledge about likely category membership (e.g., that the bug was drawn from a certain region) provides less information about the to-be-predicted feature than does the given feature (e.g., knowing that it has downswapped wings), then it seems rational to forego category information in favor of featural information (putting aside computational and memory limitations). In this view, the dominance of feature-based strategies could be seen as an active rejection of the relatively uninformative categories provided.

### Decision-Only Tasks: Implications for Category Learning and Exemplar Availability

Another characteristic feature of uncertain induction tasks, as opposed to induction tasks in which categorization is certain, concerns the relative availability of category-level and exemplar-level information during prediction. Much of the work examining

category-based induction (e.g., Rips, 1975) has used natural or social categories that are well known to participants. Moreover, in most cases, information about the category is conveyed in premises that refer to the category as a whole rather than to specific category members (e.g., “If *tigers* have a novel property X, how likely is it that this property will be shared by lions?”). In such situations, summary information about the category is readily available, but information about specific exemplars is not given; it would need to be inferred or retrieved from memory after category labels are given.

This contrasts sharply with the decision-only paradigm used in studies on induction with uncertain categories. In such paradigms, people are asked to examine the category structures but are given no explicit training in categorizing exemplars. Moreover, all category exemplars remain available for examination throughout the prediction process. Each of these factors may contribute to the dominance of feature-based over category-based induction.

The lack of explicit categorization training may mean that people form only weak representations of the typical features of each of the candidate categories and the features that distinguish these categories. There is a substantial body of evidence that shows that people form different types of conceptual representations in response to the goals of a given learning task (see Markman & Ross, 2003, for a review). In the decision-only task, there is minimal need for learning or even attending to aspects of the category structure other than those needed to answer the two key questions.

In contrast to the presumed weak encoding of category-level information, exemplar-level information is highly available in the decision-only paradigm. In conventional categorization and induction tasks, information about specific exemplars has to be retrieved from memory. There is good evidence that exemplar details become less accessible over time relative to category-level information (e.g., Bourne, Healy, Kole, & Graham, 2006; Posner & Keele, 1970). In the decision-only paradigm, however, there is no need to rely on memory for exemplars. Direct comparisons between the features of the probe instance and those of category exemplars are possible. Moreover, the online availability of category exemplars means that it is also easy to identify feature co-occurrences (e.g., in Figure 1A, the frequencies with which various wing types are paired with different types of mandibles); both of these factors favor feature-based over category-based induction.

The current studies therefore aimed to identify the key factors that lead people to use feature-based or category-based induction. Experiment 1 examined the effect of category coherence. We first describe a quantitative metric for measuring category coherence and use this to derive category structures that vary markedly in coherence. These categories were then presented in a series of decision-only induction problems. Experiment 2 moved beyond the decision-only paradigm to examine the relative impact of encoding of category-level information and exemplar availability. Participants were again presented with low- or high-coherence categories but, in this case, we varied whether or not they were trained to classify exemplars prior to inductive decisions. To examine the impact of exemplar availability at test, some of those who received classification training were also shown the categories during the induction task, whereas others had to make inferences based only on their memory for category information.

## Experiment 1

This study examined the influence of category coherence on induction with uncertain categories. To do so, it was essential to define and quantify coherence. There are relatively well-established methods for measuring category coherence in naturally occurring object and social categories, based on ratings of category uniformity, stability, and informativeness (e.g., Haslam et al., 2000). By contrast, there is no consensus over the best way to assess coherence with artificial categories (e.g., Medin, Wattenmaker, & Hampson, 1987; Rehder & Hastie, 2004). In the current experiments, we adapted an information-theoretic measure, *mutual information*, to assess (and hence manipulate) the relative coherence of categories. Mutual information,  $I(X,Y)$ , measures the amount of information provided by dimension  $X$  (e.g., the category labels) about dimension  $Y$  (e.g., a feature dimension) and is derived from the study of entropy (Shannon, 1948). Specifically, mutual information refers to the degree to which uncertainty—entropy or  $H(X)$ —in stimulus dimension  $X$  is reduced by observing dimension  $Y$ . When dimension  $X$  refers to the category label, then mutual information is a measure of category coherence, as it measures the amount of information about the features of individual exemplars that is provided by the category label. However, dimension  $X$  could equally refer to a known feature of the exemplars, in which case it would measure the amount of information about values on one stimulus dimension (e.g., type of wings) gained by knowing the value of a feature on another dimension (e.g., type of mandible). An advantage of this approach is that it allows for direct comparisons between the amount of information provided by a feature and the amount provided by a category label.

To calibrate the coherence manipulation in the present experiment, the mean mutual information provided by category labels in several previous uncertain induction tasks was calculated (these studies predominantly used artificial categories; see Appendix A). The mean information values of categories used in prior studies varied from 0.06 to 1.09. Items in our high-coherence condition were constructed so that the relevant categories had a mean information value of 0.94, which is at the upper end of the range used in previous studies (e.g., Murphy & Ross, 2005, 2010b). Low-coherence items, on the other hand, had a relatively low mean information value of 0.33 (e.g., Hayes & Newell, 2009). The psychological validity of this manipulation of category coherence was confirmed in a pilot study, which showed that participants could reliably discriminate between low- and high-coherence categories (see later text for details).

If previous findings of a strong preference for feature conjunction were a consequence of the low coherence of the categories provided, then we should again find high levels of feature-conjunction reasoning in our low-coherence condition. Increasing category coherence, however, should lead to a shift toward use of category-based induction.

## Method

**Participants.** Thirty-nine undergraduate psychology students participated for course credit. Nineteen were assigned to the high-coherence group, and the remaining 20 were assigned to the low-coherence group.

**Design and materials.** Two types of inductive problems were constructed, each involving either low- or high-coherence catego-

ries. Each group of participants received either eight low-coherence problems or eight high-coherence problems. All eight problems used the same underlying categorical structure, but the assignment of feature dimensions and feature values was randomly determined for each problem. Each induction problem involved three categories of “bugs” from a different geographical region (east, central, west), each containing 16 exemplars (see Figures 1A and 2 for examples of high- and low-coherence problems, respectively). Exemplars could take one of five values on each of five dimensions. In each problem, after studying the three categories, participants were shown a probe exemplar with a given feature (e.g., a new bug with a certain type of wings) and asked to predict its most likely category membership (classification) and its most likely value on a different feature dimension (feature prediction). Each of the eight trials consisted of a study phase, then a classification judgment, and then a feature prediction.

For the classification judgment, there were always three possible response options (three categories), whereas for feature prediction, there were five response options (five feature values). Each problem was constructed so that for a given probe, each of the three induction strategies (single-category, single-category feature-conjunction, and multiple-category feature-conjunction) clearly favored a different response in the feature prediction task (see Figure 1A for details). The remaining two response options were not favored by any strategy. The features mapped to these control response options occurred primarily outside the target category. For example, the lower two feature values shown in Figure 1B were considered control features because they were not mapped to a particular inductive strategy. Notably, these features generally had higher frequencies across all exemplars than the features mapped to the feature-conjunction inductive strategies (although this difference was quite small for the low-coherence problems). Thus, if participants were to respond solely on the basis of feature base rates, they ought to favor the control response options over the strategic responses.

For both high- and low-coherence problems, the probe exemplar was most likely (66% chance) to be a member of one category (the target category) but could belong to one other category (the non-target category). It could not belong to the third category (the irrelevant control category). The coherence manipulation is illustrated in Figures 1A and 2. The category structures used in both inductive problems are summarized in Appendix B. Figure 1B shows an example of how feature predictions were derived from the two feature-conjunction strategies and the single-category strategy (based on the high-coherence problem shown in Figure 1A). The manipulation of coherence did not affect the feature predictions shown in Figure 1B; these values were held constant across problems (the calculations in Figure 1B also apply to the low-coherence problem shown in Figure 2). Instead, coherence was manipulated by altering the degree to which each exemplar deviated from the prototypical exemplar for its category. This was subject to two constraints. First, no two exemplars were identical. Second, each exemplar (even in the low-coherence problem) resembled the prototype of the category it belonged to more closely than it resembled the prototype of either of the alternative categories.

**Pilot coherence study.** Nineteen undergraduate students were recruited to check that our manipulation of category coherence was reflected in participant perceptions of the experimental

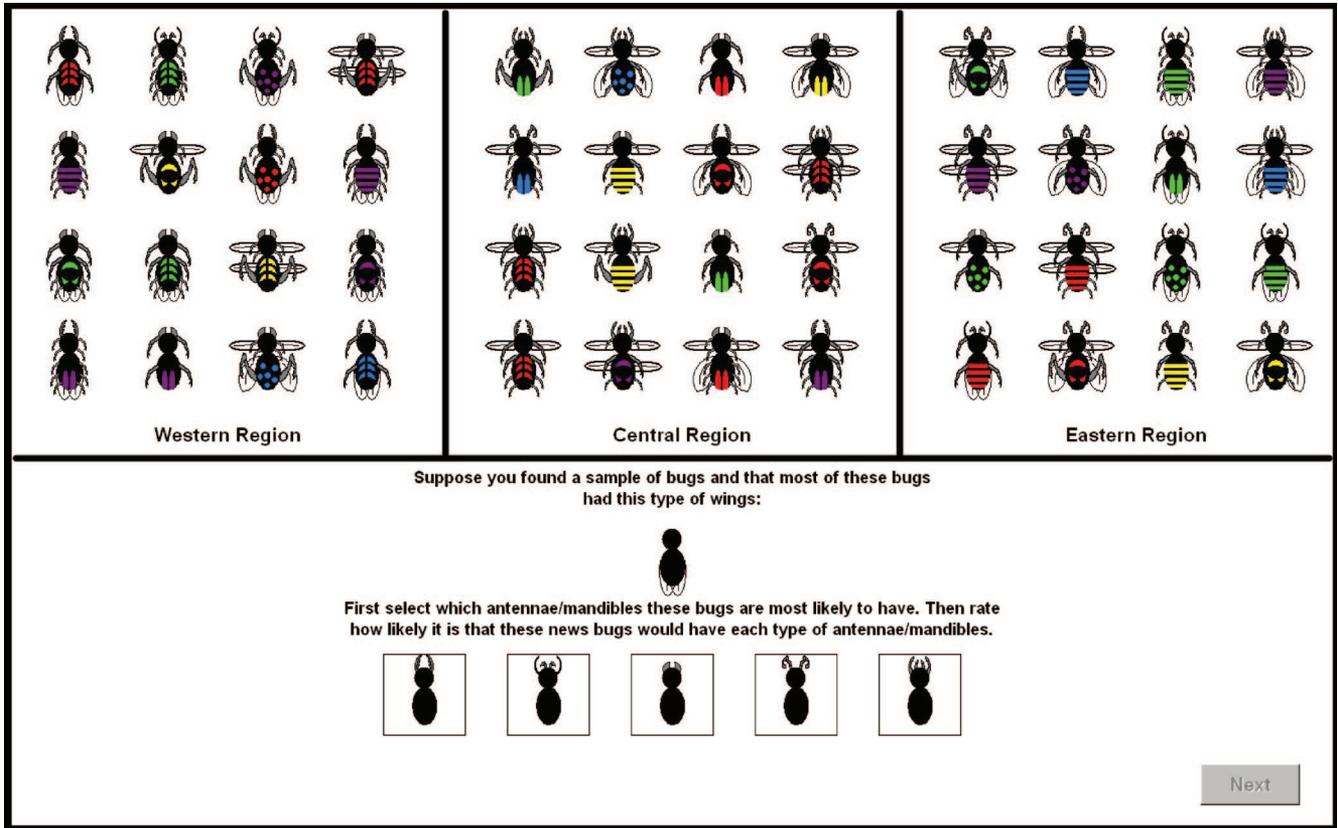


Figure 2. An example of a low-coherence uncertain induction trial used in Experiment 1.

categories. These participants were told that a professor had mixed up his samples of bugs in each of eight storage facilities. The bugs in each storage facility were originally organized into three categories, but they had been mixed together, and the professor was forced to try and sort them from memory. The participants were then told that it was their job to judge how well the professor had categorized the samples. To help them make their judgments, they were also told

- (a) if the categories are organized well, then all the bugs in a given category should look similar to each other, and (b) if the categories are organized well, then each category of bugs should look different from the other categories.

Each trial in this task commenced with a 10-s on-screen display of the 48 bugs exemplars that were to be used in training in the main experiment. The on-screen position of each exemplar was randomized, and bugs were shown without category labels or category boundaries. These exemplars then disappeared from the screen for 1 s. When the exemplars reappeared, they were arranged into three categories, in the same manner as stimuli displayed in the main experiment. Participants were asked to rate how well the professor had sorted his bugs on a 0 to 100 scale (0 = *not well at all*, 100 = *very well*). A reminder to consider all five dimensions (legs, mandibles, wings, pattern, and color) was present every time a rating was required. After the rating, the exemplars were removed from the screen, and the next trial began. There were eight

such trials, four where the bugs were shown in categories with relatively high coherence and four where they were shown in categories with relatively low coherence (these were the same low- and high-coherence category structures described earlier). Trials alternated between high- and low-coherence categories (half the participants started with a high-coherence trial, and half started with a low-coherence trial).

The mean coherence rating for the high-coherence trials was 68.94 ( $SEM = 2.73$ ), whereas the mean rating for the low-coherence trials was 37.08 ( $SEM = 3.63$ ). This difference was significant,  $t(18) = 8.00$ ,  $p < .001$ . Because the main experiment manipulated coherence between subjects, it was important to examine whether participants were sensitive to the coherence manipulation when they had observed only one category structure (of either high or low coherence). To this end, an additional analysis was carried out on participants' responses to the first category structure they were shown. For half, this was a high-coherence structure, and for half it was a low-coherence structure. The mean coherence rating (on the first trial) for those shown a high-coherence structure was 59.45 ( $SEM = 4.01$ ), whereas for those shown a low-coherence structure, it was 30.77 ( $SEM = 6.95$ ). This difference was also significant,  $t(17) = 3.94$ ,  $p < .001$ . Participants were clearly able to differentiate the coherence of the two category structures.

**Procedure.** In the main experiment, eight induction problems were presented sequentially to each participant. To begin, partic-

participants were told that they were to assume the role of an entomologist. Then, for each problem, participants were told that they would be shown pictures of bugs sampled from three different geographical regions. They were then shown an incomplete bug specimen (the probe). Their task was to first predict the region it was drawn from (classification) and then to predict a missing feature of the specimen (feature prediction). The classification judgment was always asked prior to feature prediction. This order was chosen because previous studies (e.g., Hayes & Newell, 2009; Murphy & Ross, 1994) have suggested that single-category reasoning is more likely to be observed with this question order and because this is the task order that has been used in most previous studies of induction with uncertain categories (e.g., Murphy & Ross, 1994, 2005, 2010a, 2010b).

Each problem trial began with a 30-s familiarization phase. The upper part of the screen was divided into three equal sections, one per category, and the 16 exemplars were evenly spaced within each section. Each exemplar measured 100 pixels square. Each category was labeled with a region (east, central, west) shown below the exemplars. All information, instructions, and response options were shown in the lower third of the screen. The stimuli were locked on-screen for 30 s, after which participants could progress to the first test question by clicking on an on-screen button.

The three categories were then removed, and a statement and a picture describing the given feature (e.g., a bug with wings at the rear of its body) was shown for 2 s. The classification test question was then shown for 2 s (i.e., “Which region do you think this bug came from?”). The categories then reappeared at the top of the screen, and participants indicated their classification decision by clicking on-screen buttons corresponding to the three alternative category labels. They then rated their confidence in this judgment on a 1 to 100 scale (1 = *not at all certain*, 100 = *very certain*), using the mouse to move an on-screen slider. The procedure was similar for the feature induction question (e.g., “Which type of wings do you think this bug has?”) but with one difference. After choosing the *most likely* feature prediction, they were then asked to rate the likelihoods of the four other response options on separate 0 to 10 scales (0 = *very unlikely*, 10 = *very likely*). The assignment of all stimulus variables (e.g., color, wing type, patterning) as well as the screen position of categories (target, nontarget, irrelevant) was randomly determined on each trial, with one exception. Color was considered to be of higher salience than the other dimensions; therefore, it was never the given feature dimension or the to-be-predicted feature dimension.

## Results

**Probe classification.** Two participants who were unable to identify the most likely category on the majority of trials were excluded from further analysis. For the remaining 37 participants, induction data for any trial in which the correct category was not identified were also removed. This resulted in the removal of less than 3% of the data.

Mean confidence in the classification judgment ( $M = 70.24$ ) was comparable with the conditional probability of category membership, given the provided feature value:  $p(\text{target category}|\text{given feature}) = 0.66$ . Importantly, this submaximal value suggests that participants generally saw classification of the probe as uncertain. Confidence in probe classification was slightly higher for high-

coherence ( $M = 79.45$ ) than for low-coherence items ( $M = 70.02$ ),  $t(35) = 2.74$ ,  $p = .005$ .

**Feature predictions.** The proportion of induction responses consistent with each of the three reasoning strategies (and the two nonstrategic control options) were calculated across the eight induction problems and are shown in Figure 3. Overall, feature predictions based on feature-conjunction strategies were preferred over category-based predictions. However, levels of category-based prediction were well above those based on the two nonstrategic control options, suggesting that some participants used category-based reasoning at least some of the time. The coherence manipulation did not appear to affect these general patterns. No inferential analyses were conducted on this data set, because each participant contributed more than one data point, and these data points were (a) unequal in number, (b) distributed across different cells of the analysis, and (c) not independent. Fortunately, the likelihood data described next were not subject to these limitations but showed broadly the same pattern as the forced choice data.

**Prediction likelihood ratings.** Participants’ mean likelihood ratings for each of the five possible feature prediction values, separated by group, are displayed in Figure 4. These data were analyzed in a 5 (response option: 3 strategies + 2 nonstrategic controls)  $\times$  2 (coherence) analysis of variance (ANOVA) with repeated measures on the first factor.

Four planned orthogonal contrasts were conducted. First, a main effect compared ratings for the two control response options. No significant difference was observed,  $F < 1$ . A second contrast revealed that the average of the three strategic response options (single-category, single-category feature-conjunction, and multiple-category feature-conjunction options) were rated as more likely than the average of the control response options,  $F(1, 35) = 191.30$ ,  $p < .001$ ,  $\eta_p^2 = 0.845$ . Within the strategic response options, the option associated with the category-based strategy was given significantly lower likelihood ratings than the average of the two response options associated with feature-conjunction strategies,  $F(1, 35) = 11.69$ ,  $p = .002$ ,  $\eta_p^2 = 0.250$ . No significant differences in likelihood ratings were observed between the responses associated with the two feature-conjunction strategies,  $F < 1$ . Finally, there was no main effect of coherence ( $F < 1$ ), and

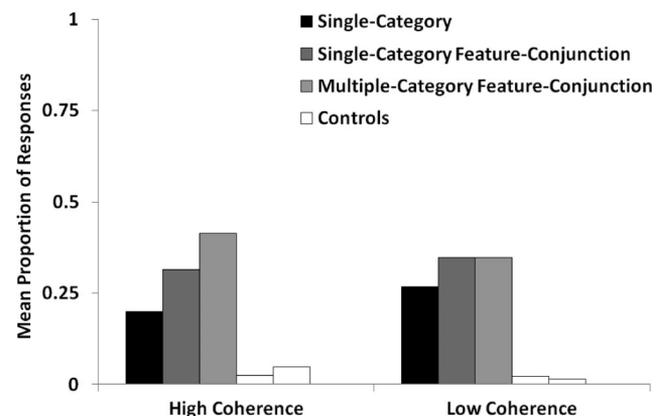


Figure 3. Mean proportion of forced choice induction responses in Experiment 1. Each column refers to the proportion of responses consistent with a particular inductive strategy.

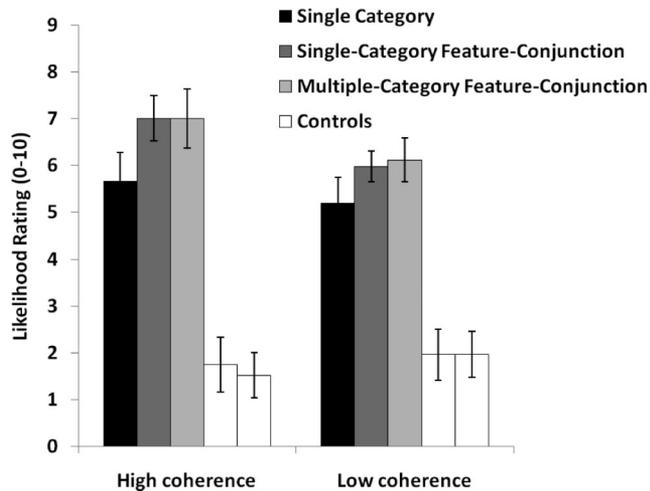


Figure 4. Mean likelihood ratings for each feature induction response in Experiment 1. Three response options were favored by a particular inductive strategy (either single-category, or single- or multiple-category feature-conjunction), whereas the remaining two acted as nonstrategic control response options. Error bars indicate standard error of the mean.

none of the strategy differences interacted with coherence, all  $F_s(1, 35) < 3.18$ .

**Consistent user analysis.** Although the inferential analyses of group likelihood data described earlier are informative, they could mask distinct behavior patterns in subsets of individuals. For example, it is possible that a minority of participants consistently chose the single-category response option but that this was masked by a majority that chose the feature-conjunction strategies.

To test this hypothesis, patterns of individual responding were examined by classifying participants according to their preferred choices on the feature prediction task. Those who predicted the feature associated with a particular strategy on the majority of trials were considered consistent users of that strategy (five or more out of a possible eight). Three people from the high-coherence group and three from the low-coherence group consistently used the single-category feature-conjunction strategy. Similarly, five people from each group used the multiple-category feature-conjunction strategy. Two people in the low-coherence group consistently used a single-category strategy, but none did in the high-coherence group. The remaining 22 participants were not classified as consistent users of any particular strategy. Because the pattern of responding was virtually identical across the category coherence manipulation, number of consistent users was summed across groups for the purposes of inferential analyses. A binomial test revealed that the number of consistent users of both single-category (six) and multiple-category (10) feature-conjunction strategies significantly exceeded chance,  $p = .03$  and  $p < .001$ , respectively. However, the number of consistent single-category strategy users (two) did not significantly exceed chance ( $p > .05$ ).

## Discussion

This study addressed one possible reason why previous work on induction with uncertain categories has found a predominance of

feature-based reasoning: the previous use of categories that lacked internal coherence. In the current study, participants made inductive judgments about categories that varied in objective coherence. Because objective category coherence has not previously been measured (or manipulated) in the manner used in the present task, it was important to first investigate whether these differences in objective category coherence were in accord with participants' subjective experience of category coherence. A pilot study confirmed that participants were sensitive to this method of manipulating coherence. However, even when more coherent category structures were used, we still found a preference for feature-conjunction strategies over category-based induction. This was apparent in both the low- and high-coherence groups, as measured by mean likelihood ratings and patterns of consistent individual responding.

Previous uncertain induction studies that examined patterns of individual responding have typically found that the majority of participants chose a particular feature-conjunction strategy and then repeatedly used this strategy on each induction task (Griffiths et al., in press; Papadopoulos et al., 2011). The present data exhibited the same pattern, as the feature-conjunction strategies were used consistently by more individuals than single-category reasoning.

However, it should also be noted that there were many more inconsistent strategy users than have been observed previously (e.g., 70% showed consistent strategy use in Griffiths et al., in press). This may reflect the greater complexity of the current categories (more feature dimensions, more feature values per dimension, more exemplars per category) compared with those used in previous work. Alternatively, it may have been due to the requirement to rate the likelihood of each possible induction response option. This could have encouraged experimentation with alternative inductive strategies over the eight inductive problems. For example, participants may have initially limited their search to the target category (relying on single-category and single-category feature-conjunction strategies), but on successive trials, they may have expanded their search to include all relevant categories, leading to multiple-category feature-conjunction reasoning. To examine this possibility, we carried out a post hoc analysis that tested whether the pattern of the forced choice responses varied across the eight induction trials. To do so, the responses consistent with each of the three induction strategies were summed separately for each of the eight trials (the two nonstrategic response options were also summed and treated as a single, additional control option). If participants systematically expanded their search set on later induction problems, then the frequency with which the various response options (e.g., single-category, multiple-category feature-conjunction) were selected should vary with the trial number. No such pattern was observed,  $\chi^2(21) = 13.45$ ,  $p = .89$ . The two feature-based strategies remained dominant throughout testing, and the relative number of inferences based on single-category conjunctions and multiple-category conjunctions did not vary across trials.

Despite this dominance, we did find two participants who appeared to consistently use single-category reasoning. Further, the likelihood rating data suggested that most participants viewed single-category induction as a more viable strategy than the nonstrategic options. These findings suggest that, although feature-based reasoning remained dominant, we did observe some level of

category-based reasoning. However, one must be cautious about this interpretation. First, the number of consistent single-category users did not exceed chance. Second, the observed moderate likelihood ratings for the single-category response option are also consistent with feature-conjunction strategy use. This is because the feature value assigned to the single-category response option also frequently co-occurred with the given feature and, thus, was a likely prediction (but not the most likely prediction) under both feature-conjunction strategies (see Figure 1B for calculations).

In sum, little clear evidence of single-category strategy use was observed in either coherence condition in Experiment 1. When faced with an inductive judgment involving uncertain category membership, participants consistently chose response options associated with the two feature-conjunction strategies and also rated these response options as most likely. This pattern of behavior did not appear to depend on the coherence of the categories involved.

## Experiment 2

Experiment 1 suggested that the use of feature-based reasoning was unlikely to be a direct consequence of the low coherence of categories used in previous work. Instead, the observed reliance on feature-based induction may be a consequence of the particular induction procedure used. Decision-only induction tasks, such as the one used in Experiment 1 (and in several other studies, e.g., Hayes & Newell, 2009; Murphy & Ross, 1994, 2010a, 2010b), expose participants to all of the relevant exemplars at the time of induction. As argued in the introduction, this procedure could promote use of feature-conjunction inductive strategies in two ways: (a) because category-level information is not encoded prior to the induction task, a suitable categorical representation may not have been learned by the time inductive judgment takes place, or (b) the availability of all exemplars throughout the inductive judgment may render any category-level representation redundant.

Importantly, both of these mechanisms could also act to undermine the influence of category coherence on inductive judgment. Although no direct effect of coherence was seen in Experiment 1, it is possible that any influence of coherence was ameliorated by, for example, a lack of experience with the categories prior to induction. Specifically, if no category representation was learned prior to the inductive decision point (or that representation was made redundant by the opportunity to inspect exemplars), then the coherence of that category representation could not be expected to influence test responding. To address this possibility, the present experiment retained the manipulation of category coherence used in Experiment 1.

In total, the present experiment manipulated three variables: prior training with the categories, exposure to the exemplars at test, and category coherence. The first two of these independent variables were manipulated in an incomplete factorial design (see Table 1). Specifically, some participants (the training and decision and the training-only groups) were provided with training with the categories prior to the critical test of inductive judgment. They were compared against groups of participants (the decision-only groups) who did not receive prior training; these decision-only groups essentially replicated Experiment 1. Comparing the pattern of feature inductions made by the trained and nontrained groups reveals whether an opportunity to learn a category representation affects subsequent inductive strategy use. We predicted that prior

Table 1  
*Design of Experiment 2*

Categories shown during training?	Categories shown during induction decision?	
	Yes	No
Yes	Training and decision group	Training-only group
No	Decision-only group	—

*Note.* Prior training with categories and the presence of exemplars at test were manipulated to produce an incomplete factorial design. The name of each group refers to when the category stimuli were present. The decision-only group replicates Experiment 1. A third independent variable, category coherence, was manipulated orthogonally to the two shown here.

training with the categories would facilitate development of a suitable category-level representation, which would in turn promote use of category-based inductive strategies at test.

Of the groups that received training prior to the induction test, half were shown the category exemplars during test (the training and decision groups), and half were not (training-only groups). If the presence of the exemplars at test rendered any learned categorical representation redundant, then the training and decision groups should exhibit a stronger reliance than the training-only groups on feature-conjunction inductive strategies (rather than on category-based induction). Finally, the coherence of the categories that were trained and tested was manipulated orthogonally to the other independent variables. Comparison of the high- and low-coherence groups assesses whether inductive strategy choice depends on category coherence.

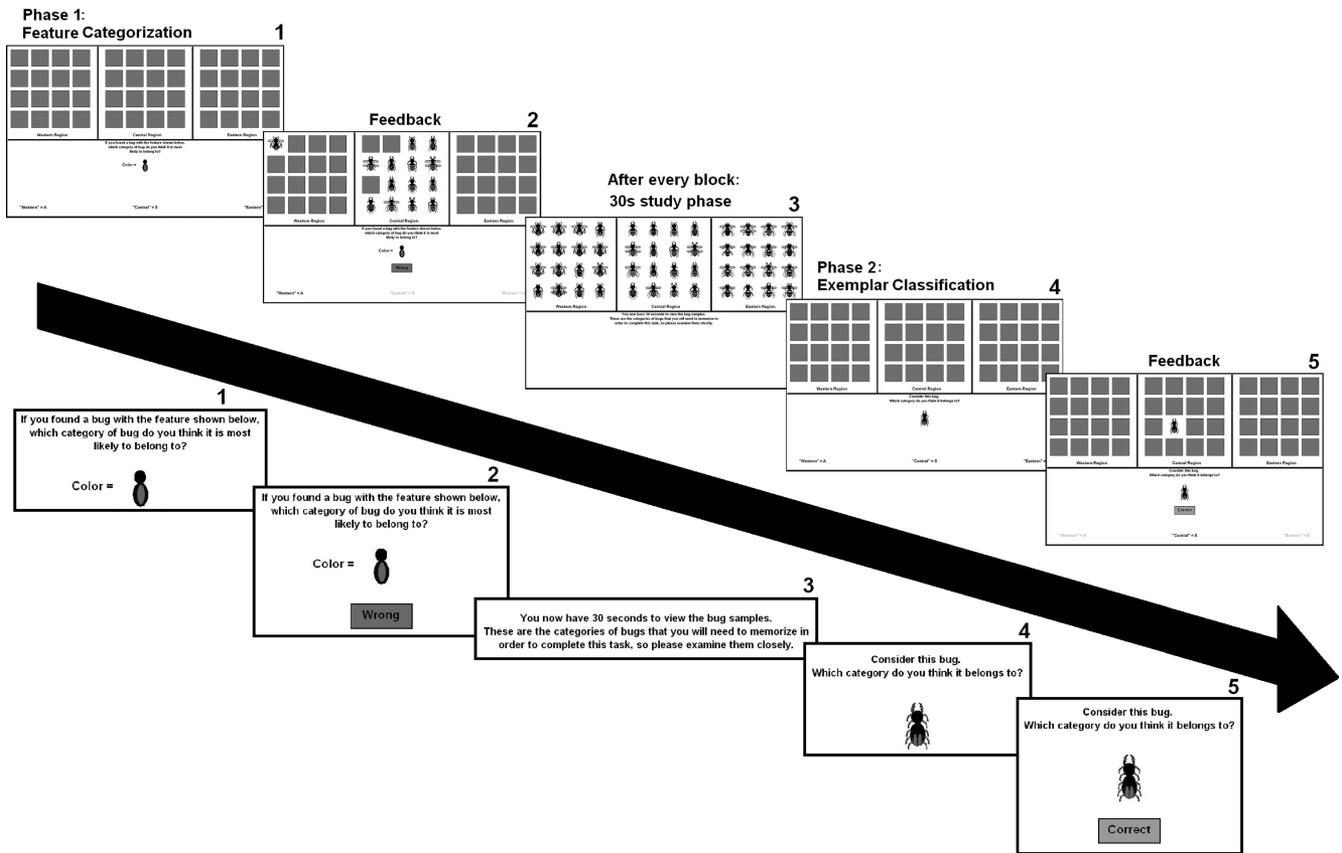
## Method

**Participants.** One hundred and eight undergraduate psychology students participated in exchange for course credit.

**Procedure.** All groups were tested with the same feature induction task used in Experiment 1. Half were shown the same high-coherence categories as were used in Experiment 1, and half were shown the low-coherence categories.

Those in the decision-only condition were not given any training and essentially replicated the procedure for Experiment 1. They were given the same instructions as in Experiment 1 and proceeded directly to the induction test. The test procedure was the same as in Experiment 1: Participants were given eight induction problems, each using different instantiations of the underlying category structure. In contrast, those in the training-only and training and decision groups (four groups in total) were first given extensive training with the categories (broken into two phases), before being given a single inductive problem at test. This single problem referred to the categories on which these participants had been trained. Hence, unlike the decision-only groups, the trained groups received only a single inductive problem at test.

To begin, the four trained groups were asked to assume the role of an entomologist and learn how to classify different types of bugs. They were then given extensive training with the three bug categories. This training was divided into two phases (screenshots from both phases are shown in Figure 5). Each phase began with a study period in which all members of each of the three bug categories were visible on-screen for 30 s (see Panel 3 of Figure 5).



*Figure 5.* Depiction of procedure and stimulus layout during training in Experiment 2. In Phase 1, a single feature value is shown, and participants are required to identify the category to which bugs with that feature value are most likely to belong (Panel 1). Feedback is provided on each trial (Panel 2). In Phase 2, a single complete exemplar is shown, and participants are asked to categorize that exemplar (Panel 4). Feedback is provided after every judgment (Panel 5). Study trials (Panel 3) are shown at the beginning of each phase and every so often during the phases (after every 15 trials in Phase 1 and after every 48 trials in Phase 2). The text shown on-screen in each panel is expanded below the arrow. Note that the decision-only groups were not given this training.

After this initial study phase, trial-by-trial classification learning commenced. On each trial in the first training phase, a probe exemplar with one feature (e.g., downswept wings) was shown, and people were asked to choose the category to which the exemplar was most likely to belong. During this classification process, the bug images were removed from the screen, but the lines separating the categories of bugs (eastern region, etc.) remained on the screen. After making a prediction, corrective feedback was provided (*correct* or *wrong* was shown on-screen), and all of the exemplars with the given feature were revealed. Note that because features were probabilistically (not deterministically) related to category membership, this meant that exemplars from the incorrect categories were often also revealed. This emphasized the probabilistic relationship between the feature values and the categories.

Throughout the first training phase, people were asked to classify only items based on the presentation of typical features (see Panels 1 and 2 of Figure 5). This meant that only three values per dimension were used (a total of 15 features). The

initial trials were blocked by dimension to aid learning. This meant, for example, that a participant was first asked to classify a series of exemplars, each of which had one of the three types of wings that were typical of the three training categories. They then proceeded to exemplars with typical values on the leg dimension and so on for all five dimensions. The order of presentation of the dimensions was randomized, with the constraint that color was always shown first. Two further 30-s study phases were shown after the third and fifth dimensions (after the ninth and 15th trials) to remind participants about the structure of each category.

After the structured initial trials, participants were presented with blocks of 15 trials in which these same incomplete exemplars (each with one feature that was typical of a particular category) were shown in a random order. Each block began with a 30-s study phase. Participants were shown between two and seven additional training blocks, depending on their classification accuracy. When a participant accurately classified 14 or more exemplars (93.33%) in a single training block, they progressed immediately to Phase 2

of the training. If they did not do so within seven blocks, they were removed from the experiment.

On each trial in the second training phase, participants were shown a single complete exemplar and asked to categorize it (see Panels 4 and 5 of Figure 5). After each classification judgment, corrective feedback was provided (*correct* or *wrong* was displayed), and the correct category assignment of the item was displayed. Phase 2 training was organized into blocks, in which each of the 48 exemplars was shown once in a random order. Each block began with a 30-s study phase. Training continued until the participant recorded 24 or more correct classifications of the exemplars drawn from the target and nontarget categories (75% accuracy). Knowledge of the control category exemplars was not necessary for the final inductive judgment, so they were excluded from the performance criterion. If the criterion was not reached in four training blocks, the participant was removed from the experiment.

All participants then completed the same induction test as that used in Experiment 1. For those in the training (training and decision, training-only) groups, this test immediately followed the completion of training. Participants were shown a probe exemplar and were first asked to classify that exemplar and then predict a missing feature. Note, however, that those in the decision-only condition received eight induction problems (each based on a different set of categories), whereas those in the training groups received only one (based on the categories with which they had been trained). The induction problem that was common to all groups was structurally identical, although the type of the given feature that was presented for this problem (e.g., which type of legs, wings, etc.) was randomly determined for each participant.

Those in the decision-only and training and decision groups were shown the full set of exemplars for each category during testing (at the top of the screen), but those in the training-only condition saw only the probe exemplar. Note also that after participants classified the probe exemplar, they were asked to rate the likelihood of all three possible category assignments on separate 0 to 10 response scales.

## Results

**Training data.** All participants who were required to learn the high-coherence category structures reached criterion performance and proceeded to the induction test, whereas only 44% of those given low-coherence structures reached criterion (we consider the performance of the excluded participants later). For the high-coherence categories, the mean number of blocks to criterion was 3.2 in Phase 1 (with possible values of two to seven blocks) and 1.24 in Phase 2 (with possible values of one to four blocks).<sup>1</sup> In contrast, of the participants who reached criterion in the low-coherence groups, the mean number of blocks to criterion was 4.5 in Phase 1 and 1.57 in Phase 2.

**Probe classification.** Thirty-four people did not correctly classify the probe exemplar at test (or on the first test trial, for the decision-only groups) and were thus excluded from the analysis. Fifteen of these participants were from the training-only groups, seven were from the training and decision groups, and 12 were from the decision-only groups. These exclusion rates did not significantly differ between groups,  $\chi^2(2) = 3.00, p > .05$ . Where possible, these participants were replaced to yield approximately

20 people per group (there were 21 people in both the high-coherence decision-only and training-only groups, 20 in the high-coherence training and decision group, and 23 in the low-coherence decision-only group). Because relatively few people reached criterion in the low-coherence training groups, these groups were smaller: There were 10 people in the low-coherence training-only group and 13 in the training and decision group.

In the comparison of performance between the decision-only and other groups on the classification and induction questions, only the data from the first induction problem presented to the decision group were analyzed (data for the remaining problems are presented in the next section). As expected, averaged across the groups, the mean likelihood ratings for the target category ( $M = 8.81$ ) were higher than for the average of the nontarget ( $M = 4.17$ ) and irrelevant categories ( $M = 1.32$ ), all  $F_s(1, 102) = 573.61, p < .001, \eta_p^2 = 0.849$ . This preference for the target category was not affected by any of the independent variables (training, presence of the exemplars at test, or category coherence), all  $F_s < 1.59$ . Overall, participants also gave higher likelihood ratings for the plausible nontarget category than for the irrelevant control category,  $F(1, 102) = 71.45, p < .001, \eta_p^2 = 0.412$ . This preference, however, was significantly larger for participants in the decision-only groups than for those given training,  $F(1, 102) = 31.02, p < .001, \eta_p^2 = 0.233$ . Similarly, of the groups given training, those participants who were shown the exemplars at test were better able to discriminate between the plausible alternative (nontarget) category and the irrelevant control category,  $F(1, 102) = 6.76, p = .011, \eta_p^2 = 0.006$ , than those in the training-only condition.

**Feature predictions.** To compare induction across all six groups, responses on the first induction test problem shown to the decision-only group were compared against responses to the single induction test problem given to those in the two trained groups. The number of participants who chose the feature associated with each reasoning strategy on the first induction problem was tallied (see Figure 6). These data were entered into three (orthogonal) chi-square tests to assess the influence of the three independent variables (training, presence of exemplars at test, and coherence) on people's use of inductive strategies. All analyses were performed on data that were normalized with respect to cell size, to avoid biases due to unequal cell sizes. Because of this adjustment, the values entered into these analyses differ slightly from those shown in Figure 6. To aid with interpretation of the nonparametric analyses, the data pertinent to each of the individual analyses are plotted in Figure 7.

The first analysis pooled all participants who received training and compared them against those who did not. A significant difference in response choices was observed,  $\chi^2(3) = 12.05, p < .01$ . As can be seen in the upper panel of Figure 7, participants given training were most likely to use a category-based strategy, whereas those not given training were most likely to use the multiple-category feature-conjunction strategy. A second chi-square analysis examined whether there was any effect of provid-

<sup>1</sup> Although the high-coherence groups reached the 75% criterion in 1.24 blocks, on average, they proceeded to the test phase only after they reached 93% performance (which all high-coherence participants did). This means that the high-coherence groups received 2.26 training blocks, on average, before test.

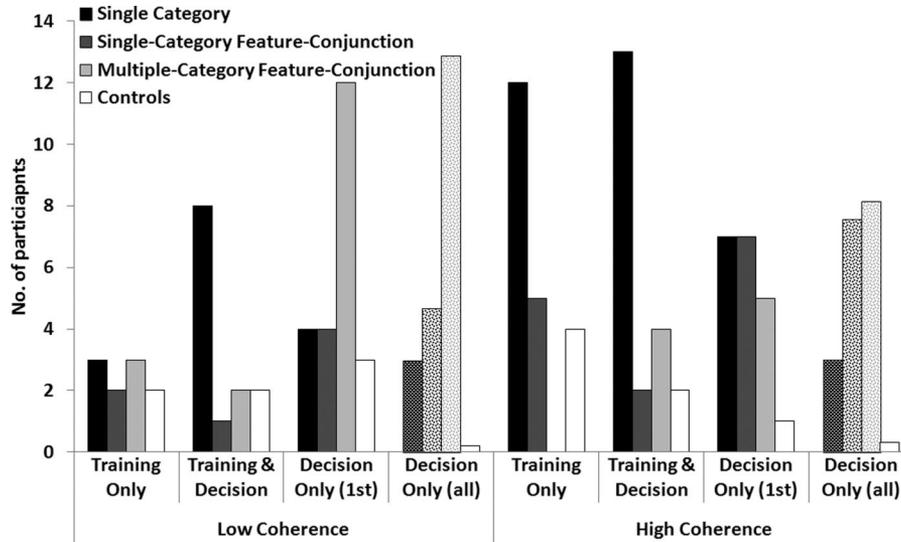


Figure 6. Frequency of forced choice responses in Experiment 2, divided by group. Each column refers to the number of participants who selected the response option consistent with a particular induction strategy. For the decision-only groups, the responses on the first induction task (of eight) are plotted in the left-hand (full) bars, and the normalized average of their responses across all eight induction problems is plotted in the right-hand (dotted) bars.

ing the categories on-screen during judgment. This was done by pooling the data from the training-only groups and comparing them against the pooled data of the training and decision groups (the decision-only groups' data were excluded). No significant difference in response choice was observed between these conditions,  $\chi^2(3) = 3.05$ ,  $p > .05$ . A final analysis examined whether coherence affected response choice. To this end, data were pooled across all the groups shown high-coherence categories, and these data were compared against those shown low-coherence categories. A marginally significant difference was observed,  $\chi^2(3) = 6.40$ ,  $p = .09$ . If anything, it appears that those given the high-coherence categories showed a stronger bias toward the single-category strategy, whereas those given low-coherence categories showed an equivalent bias for the multiple-category feature-conjunction strategy and the single-category strategy.

As noted earlier, the decision-only groups made responses to multiple induction problems at test. To check that the patterns of inductive prediction across the eight items for this group were similar to those seen in Experiment 1, the same analysis of consistent individual responding that was used in Experiment 1 was also conducted on these data. The relevant data are shown in the right-hand (dotted) bars of Figure 6. The figure shows that, as in Experiment 1, feature-based predictions again dominated responding in these groups. Twenty people consistently used the multiple-category feature-conjunction strategy (14 from the low-coherence group), and six consistently used the single-category feature-conjunction strategy (two from the low-coherence group). In contrast, only two people consistently used the category-based strategy (both were in the low-coherence group). Overall, the frequency of both kinds of feature-based induction was significantly greater than chance on a binomial test, highest  $p < .01$ , but the number of consistent category-based strategy users did not exceed chance,  $p > .05$ . Interestingly, the pattern of consistent

users significantly differed between the two coherence groups,  $\chi^2(3) = 8.04$ ,  $p < .05$ , whereby the low-coherence group contained a disproportionate number of multiple-category feature-conjunction users relative to the high-coherence group.

**Prediction likelihood ratings.** The mean likelihood ratings for each of the induction response options are depicted in Figure 8. As in the previous analysis, the data from the decision-only group are taken from their first induction task (the solid bars). The likelihood ratings reveal a pattern similar to the one for the forced choice induction responses. To analyze these data, we carried out a 3 (group: decision only, training only, training and decision)  $\times$  5 (strategy: 3 induction strategies + 2 nonstrategic controls)  $\times$  2 (coherence) ANOVA with strategy treated as a repeated measures factor. Planned orthogonal contrasts were used to examine specific effects based on the group and strategy factors.

A main effect was observed whereby the average of the three strategic response options was rated as more likely than the average of the two nonstrategic controls,  $F(1, 102) = 108.03$ ,  $p < .001$ ,  $\eta_p^2 = 0.514$ . This difference between the rated likelihood of the strategic and nonstrategic controls was larger in the decision-only groups than the trained groups,  $F(1, 102) = 10.31$ ,  $p = .002$ ,  $\eta_p^2 = 0.101$ . As expected, there was no significant difference between the ratings given to the two control response options across groups,  $F(1, 102) = 2.96$ ,  $p > .05$ . A second contrast revealed that the single-category response was generally rated as more likely than the two feature-conjunction strategies,  $F(1, 102) = 11.60$ ,  $p = .001$ ,  $\eta_p^2 = 0.102$ . Critically, however, this difference was significantly larger for the two groups given classification training than it was for the decision-only group,  $F(1, 102) = 16.78$ ,  $p < .001$ ,  $\eta_p^2 = 0.141$ . Simple effects contrasts were conducted to examine this interaction. The single-category strategic option was rated higher than the average of the feature-conjunction strategy options in the training-only group,  $F(1, 29) = 13.09$ ,  $p = .001$ ,  $\eta_p^2 = 0.311$ , and

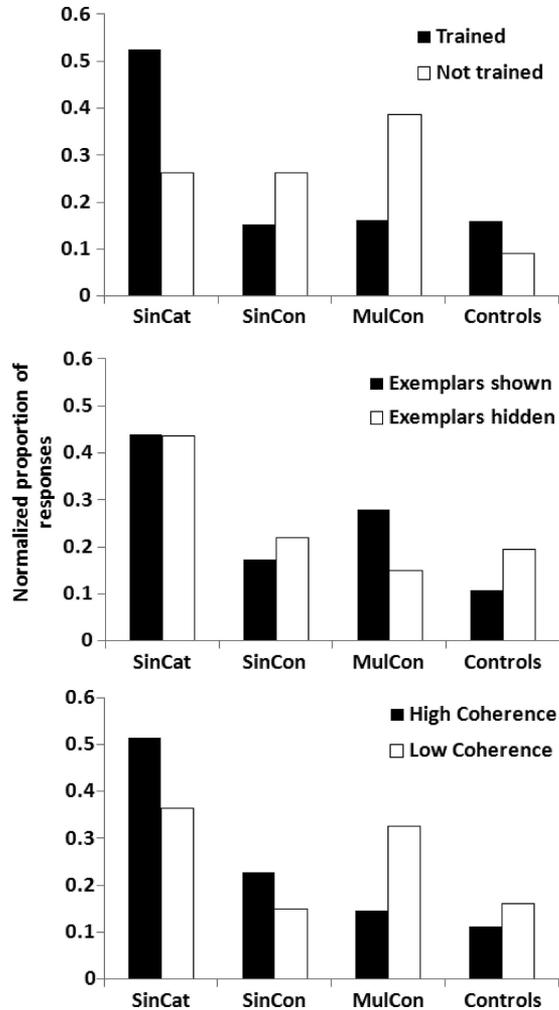


Figure 7. Normalized forced choice data that were used in the nonparametric analyses in Experiment 2. The top panel compares the pooled pattern of response choices for the trained and untrained groups. The middle panel compares the trained participants who were also shown the exemplars at test with the trained participants who were not. The lower panel compares the forced choice behavior of those given high-coherence categories with those given low-coherence categories. SinCat = single-category; SinCon = single-category feature-conjunction; MulCon = multiple-category feature-conjunction; Controls = the average of the two nonstrategic response options.

the training and decision group,  $F(1, 31) = 10.12, p = .003, \eta_p^2 = 0.246$ , but not in the decision-only group,  $F(1, 42) = 1.77, p > .05$ . There were no differences between the two training conditions (training only, training and decision) in the rated likelihood of single-category relative to the feature-conjunction options,  $F < 1$ . No main-effect differences were seen between the ratings given to the single- and multiple-category feature-conjunction strategic options,  $F < 1$ , and no interactions with group membership, all  $F_s(1, 102) < 1.11$ . Category coherence did not interact with any of the contrasts of interest, all  $F_s < 1$ .

**Excluded participants.** Approximately 56% of those trained with the low-coherence categories did not reach the 75% accuracy

criterion. Once we realized that this would be a large portion of our participant pool and that the performance of the excluded people might be interesting, we started collecting test data from these excluded participants. For this reason, our data collection from the excluded groups is partial. Nevertheless, the excluded participants from the training and decision group ( $n = 9$ ) form an interesting control for the included participants from this group. This is because both the included and excluded subgroups were given the same training and were provided with the exemplars during the induction test. However, one could assume that only the included subgroup had an appropriate category representation at the time of induction. If this is the critical determinant of category-based induction, then the included participants should show single-category induction, whereas the excluded participants should use feature conjunction. As can be seen from the mean induction likelihood ratings of the two groups, summarized in Figure 9, this is precisely what occurred.

A 2 (included/excluded)  $\times$  5 (response option) ANOVA on the likelihood ratings from the included and excluded individuals confirmed that, averaged across the two subgroups, there was no difference in the mean likelihood rating given to the single-category option and the two feature-conjunction response options,  $F < 1$ . However, there was a significant interaction between inclusion status and response type,  $F(1, 20) = 13.76, p = .001, \eta_p^2 = 0.408$ . Simple-effects analyses show that the excluded participants rated the feature-conjunction responses as more likely than the category-based response,  $F(1, 8) = 6.05, p = .039, \eta_p^2 = 0.431$ , whereas the included participants showed the opposite pattern,  $F(1, 12) = 8.28, p = .014, \eta_p^2 = 0.408$ . Clearly, category learning shaped the feature inductions participants made at test.

## Discussion

The decision-only groups in this study most closely matched the procedure used to study induction with uncertain categories in Experiment 1 and in previous work in this field (e.g., Murphy & Ross, 1994). As in the previous studies, our decision-only participants showed a preference for feature-based (single- and multiple-category feature-conjunction reasoning) over category-based reasoning. Several participants in these groups consistently used each of the feature-conjunction strategies, but only two consistently used a single-category strategy (which was not significantly greater than the number expected by chance). Providing participants with extensive classification training before induction and removing the categories from the screen while predictions were being made (the training-only groups), however, caused a reversal of this preference. The training-only groups showed a clear preference for inductive reasoning based on the typical features of the target category (single-category reasoning) over both kinds of feature conjunction.

To isolate the cause of this switch from feature-conjunction to category-based responding, a third condition was tested. The training and decision groups received training with the categories prior to induction and were also provided with the categories during the critical induction task. These groups also showed a preference for the single-category strategy over the feature-conjunction strategies. This suggests that the main factor driving the choice of inductive strategy is prior training with the categories, rather than the availability of exemplar information during induction. This

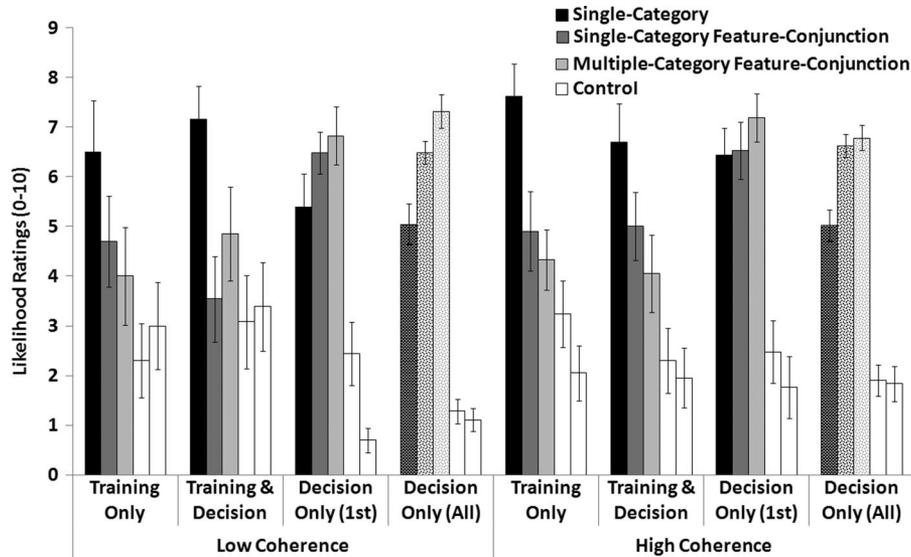


Figure 8. Mean likelihood ratings for each feature induction response option in Experiment 2. For the decision-only group, the responses on the first induction task (of eight) are plotted on the left (solid bars), and their mean performance across all eight problems is plotted on the right (dotted bars). Error bars indicate standard error of the mean.

was confirmed by the follow-up analyses of the included/excluded portions of the low-coherence training and decision group, in that only those participants who learned the training categories showed evidence of category-based induction at test.

In contrast with Experiment 1, the feature prediction data in Experiment 2 suggest that category coherence may play a small direct role in guiding inductive strategy selection. Specifically, a trend was observed whereby those given high-coherence categories more often selected the single-category response option than the multiple-category feature-conjunction option, whereas those given the low-coherence categories did not show a preference between these options. Furthermore, in the decision-only condition, fewer participants in the high-coherence group consistently

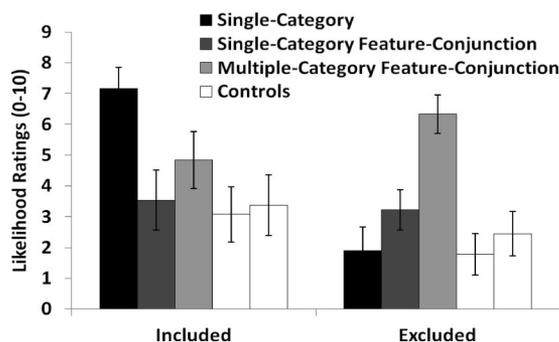


Figure 9. Mean feature induction likelihood ratings of the included and excluded portions of the low-coherence training and decision group in Experiment 2. The participants who did not learn to categorize the exemplars in training (excluded subgroup) predominantly used a multiple-category feature-conjunction strategy, whereas those who could categorize the exemplars (included subgroup) favored the category-based inductive strategy. Error bars indicate standard error of the mean.

used multiple-category feature-conjunction than in the low-coherence group. Both trends are consistent with the idea that category-based reasoning is more likely to be used when category labels are informative as to the likely features of category members (i.e., when coherence is high). Such a pattern could be the result of strategic use of category information, whereby category membership was only utilized if it provided more predictive power than the observed feature values. However, this interpretation must be treated cautiously, because the first trend did not reach significance, and the second was observed only in the present data (and not in Experiment 1).

It is clear, however, that coherence can have an indirect effect on inductive strategy use. All of the participants given the high-coherence categories reached the performance criterion during the learning task, whereas the majority (56%) given the low-coherence categories did not reach criterion. Further, only those participants who learned the relevant category structures in training showed evidence of single-category inductive reasoning at test. Thus, overall, participants given a high-coherence structure were more likely to adopt a category-based strategy at test, presumably because they had encoded the relevant conceptual representation during training and were consequently able to implement it at test.

## General Discussion

The current experiments examined why many studies of induction with uncertain categories have found a dominance of feature-based reasoning over category-based reasoning, whereas category-based reasoning is prevalent in other kinds of inductive judgments. Two possible sources of this difference were the previous use of categories that lacked internal coherence and the use of a decision-only induction procedure, which minimizes the need to learn and remember category boundaries. Experiment 1 addressed the first

hypothesis by manipulating category coherence in the standard, decision-only procedure. This experiment found no influence of category coherence on inductive judgment; feature-conjunction reasoning remained the dominant inductive strategy.

Experiment 2 extended these findings by simultaneously manipulating coherence as well as two elements of the standard induction procedure: experience with the categories prior to induction and on-line availability of category exemplars during the induction test. Specifically, participants (a) were required to learn the categories prior to induction or not, (b) were provided with the exemplars on-screen during induction or not, and (c) were shown either high- or low-coherence categories. The experiment revealed that the critical factor in determining inductive strategy was the level of experience with the categories prior to induction: Those given extensive training with the categories tended to use a category-based inductive strategy, whereas those not given training tended to use feature-conjunction strategies. In addition, some evidence was found to suggest that participants were more likely to use a category-based strategy when category coherence was high. However, the strongest effect of coherence on inductive reasoning was indirect. That is, high-coherence categories were easier to learn, which in turn encouraged use of category-based inductive reasoning at test.

These findings are important because they help to resolve the issue of when people will use salient exemplar features as opposed to category-level information as a basis for feature induction. The key difference between the present decision-only task and other inductive judgment tasks appears to be that the decision-only task does not require participants to encode the typical features of the candidate categories. Other induction tasks either train participants with the categories through trial-by-trial querying of category labels or missing features (e.g., Chin-Parker & Ross, 2002) or use highly familiar categories, whose typical features are well known (e.g., Rips, 1975). This means that category-level information is entrenched and highly available for subsequent predictions.

These data are broadly consistent with the view that the specific goals of a given categorization task drive the type of conceptual representation formed (Markman & Ross, 2003). In decision-only induction problems, participants are not required to engage in the demanding process of encoding the feature structures of the candidate categories. Therefore, when an unambiguous prediction can be made on the basis of feature co-occurrences (as in Experiment 1) the majority of participants are likely to make a feature-based prediction. Our data show, however, that when the category representation is strengthened through appropriate learning, category-based induction overrides feature-based induction in a majority of cases.

### **Relationship to Previous Studies of Category Learning and Induction With Uncertain Categories**

It should be noted that although most previous studies of induction with uncertain categories have used a decision-only paradigm, a handful have involved some form of classification training before inductive probes were presented. Murphy and Ross (1994) briefly reported a study (Experiment 3) where participants were required to memorize individual exemplars along with their category membership, before predicting unknown features of probe items. The results were identical to previous decision-only studies

involving the same stimulus sets; in both cases, feature predictions were based on consideration of only the target category. Verde, Murphy, and Ross (2005) trained participants to categorize exemplars before presenting them with a speeded induction test in which a feature of a probe item was presented and participants had to predict the feature value on another dimension. Under this speeded induction test, Verde et al. reported evidence suggesting that people may have been using multiple-category reasoning to make predictions.

The most important point to note here is that unlike the current experiments, neither of these previous studies attempted to discriminate between category-based and feature-based reasoning (with their main focus being on the use of single vs. multiple categories). In Murphy and Ross (1994), the predictions of single-category reasoning and single-category feature-conjunction reasoning were confounded, with both approaches leading to the same predicted feature (multiple-category conjunction reasoning did not lead to a clear feature prediction in this study and so was unlikely to have been used). Similarly, in Verde et al.'s (2005) studies, the apparent evidence for multiple-category reasoning is likely to have been produced by participants engaging in multiple-category feature-conjunction reasoning (Newell et al., 2010). Thus, the primary contribution of the present study (Experiment 2, in particular) is that it is the first to examine the influence of prior category training on the relative adoption of category-based and feature-based inductive strategies.

### **Implications for Theories of Category Learning and Induction**

How then is a conceptual representation that is useful for induction formed, and what format might it take? One way to address this question is to consider how our data might be explained by existing models of category learning. It is important to note, however, that the current studies were not designed to discriminate between competing models.

The first phase of categorization training in Experiment 2 involved learning to associate the typical feature values for each category with the appropriate category labels. Such training might be expected to induce a summary-level representation of each category (e.g., "most bugs from the western region have eight legs"). In contrast, the second phase of training involved classification of complete exemplars and, thus, may have led to categories being encoded as sets of specific exemplars (e.g., Kruschke, 1992; Nosofsky, 1986).

Exemplar-based models (e.g., Kruschke, 1992; Nosofsky, 1986) seem to provide a good account of the present data. For example, Nosofsky's (1986) generalized context model considers feature induction to consist of two processes. First, the agent calculates the similarity of the probe exemplar to the various exemplars that have been encountered and encoded (weighted by attention). Next, the most similar exemplars are used to calculate the most likely value of the missing feature of the probe exemplar. Suppose that the difference between the groups given training and those not given training in the present experiments is that the trained participants have encoded the category labels of the encountered exemplars, whereas those who have not been trained have not. Under this assumption, for the nontrained groups (both groups of Experiment 1 and the decision-only group of Experiment 2), the

most similar exemplars to the probe exemplar are those with the given feature, and thus, induction ought to be based on these exemplars. Therefore, in these circumstances, exemplar-based models predict feature-conjunction strategy use (as observed).

After training, however, each exemplar possesses one additional feature: its category label. It has been suggested that category labels can be considered as a stimulus feature with high attentional weight (e.g., Anderson, 1991; but see also Yamauchi & Markman, 2000). If this attentional weight is high enough, the category label feature may completely overshadow the given feature (particularly because participants are first asked to classify the probe exemplar). This would lead to the probe exemplar having greater similarity to all exemplars from the target category than to any other exemplars (irrespective of their possession of the given feature). It is under these conditions that single-category reasoning would be expected. This prediction is consistent with the trained groups' use of single-category reasoning in Experiment 2. Interestingly, if the overshadowing of the given feature by the category label feature was only partial, then this would lead to single-category feature-conjunction reasoning; if there was no overshadowing, it would lead to multiple-category feature-conjunction reasoning. Thus, with a few enabling assumptions, standard exemplar models can account for most aspects of the present data.

Anderson's (1991) rational model (identified earlier with the multiple-category inductive strategy) can also account for the pattern of responding of the trained groups in Experiment 2. Like the generalized context model, the rational model retains information about individual exemplars but uses a different method for generating feature inferences. Anderson's (1991) model first uses the given feature of the probe exemplar to identify a subset of likely classifications. It then discards this given feature and instead uses the typical feature values of each of the possible classifications (weighted by the likelihood of each classification) to form its prediction. The present category structures were not designed to directly assess this multiple-category account, but on examination this strategy leads to the same feature prediction as the single-category inductive strategy (because of the strong family resemblance structure of the present categories). Thus, the trained groups' preference for category-based induction in Experiment 2 is also consistent with Anderson's model.

Further, one could argue that this model explains the use of multiple-category feature conjunction by the decision-only groups (in Experiments 1 and 2). In the absence of training, this model suggests that participants will not learn to cluster the exemplars appropriately. Consequently, participants may treat every exemplar as its own cluster. Such an approach would result in inductive predictions consistent with the multiple-category feature-conjunction strategy, as seen in the decision-only groups.

However, it is not clear how Anderson's (1991) model explains the consistent use of single-category feature conjunction by many participants in Experiments 1 and 2. Under this model, people should not limit their search to exemplars with the given feature in a single category; they ought to either consider all exemplars with that feature (prior to training; multiple-category feature-conjunction reasoning) or consider all exemplars in the target category, with or without the given feature (after training; single-category reasoning).

## Conclusion

The present experiments found that prior training with a category leads to greater use of category-based reasoning and less use of feature-based reasoning in uncertain induction tasks. We observed some suggestive evidence that category coherence can directly influence inductive strategy choice. More important, there was a strong indirect effect of category coherence on induction, in that high-coherence categories were easier to encode, which in turn increased the likelihood of category-based induction. No evidence was found to suggest that the presence of the categories on-screen during induction influenced inductive reasoning.

Together, these observations suggest that the reason that feature-based inductive reasoning has often been favored in uncertain induction tasks, whereas category-based inductive reasoning is prevalent in more traditional induction tasks, centers on the level of prior experience with the provided categories. When participants have a lot of experience with candidate categories, and these categories are coherent, they have an opportunity to learn about category boundaries and the distribution of features within categories. Consequently, they are more likely to use category-based rather than feature-based induction. Returning to our opening example, when confronted with a colleague who supports a woman's right to abortion but opposes government-subsidized health care, an avid follower of American politics may try to leverage their knowledge of political parties (i.e., the colleague may belong to the secular Republican category) to predict their attitude toward same-sex marriage, whereas a less politically savvy person may base their prediction on the specific opinions (i.e., features) that their colleague possesses.

## References

- Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429. doi:10.1037/0033-295X.98.3.409
- Bourne, L. E., Healy, A. F., Kole, J. A., & Graham, S. M. (2006). Strategy shifts in classification skill acquisition: Does memory retrieval dominate rule use? *Memory & Cognition*, *34*, 903–913. doi:10.3758/BF03193436
- Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition*, *30*, 353–362. doi:10.3758/BF03194936
- Griffiths, O., Hayes, B. K., & Newell, B. R. (in press). Where to look first for an explanation of induction with uncertain categories. *Psychonomic Bulletin and Review*.
- Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, *65*, 137–165. doi:10.1016/S0010-0277(97)00042-5
- Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology*, *39*, 127–139. doi:10.1348/014466600164363
- Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 278–292.
- Hayes, B. K., Kurniawan, H., & Newell, B. R. (2011). Rich in vitamin C or just a convenient snack? Multiple-category reasoning with cross-classified foods. *Memory & Cognition*, *39*, 92–106. doi:10.3758/s13421-010-0022-7
- Hayes, B. K., & Newell, B. R. (2009). Induction with uncertain categories: When do people consider the category alternatives? *Memory & Cognition*, *37*, 730–743. doi:10.3758/MC.37.6.730
- 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. 209–214). Mahwah, NJ: Erlbaum.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44. doi:10.1037/0033-295X.99.1.22
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–613. doi:10.1037/0033-2909.129.4.592
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, *19*, 242–279. doi:10.1016/0010-0285(87)90012-0
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–315. doi:10.1037/0033-295X.92.3.289
- Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categories. *Cognitive Psychology*, *27*, 148–193. doi:10.1006/cogp.1994.1015
- 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, United Kingdom: Cambridge University Press.
- Murphy, G. L., & Ross, B. H. (2010a). Category vs. object knowledge in category-based induction. *Journal of Memory & Language*, *63*, 1–17. doi:10.1016/j.jml.2009.12.002
- Murphy, G. L., & Ross, B. H. (2010b). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 263–276. doi:10.1037/a0018685
- Newell, B. R., Paton, H., Hayes, B. K., & Griffiths, O. (2010). Speeded induction under uncertainty: The influence of multiple categories and feature conjunctions. *Psychonomic Bulletin & Review*, *17*, 869–874. doi:10.3758/PBR.17.6.869
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57. doi:10.1037/0096-3445.115.1.39
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200. doi:10.1037/0033-295X.97.2.185
- Papadopoulos, C., Hayes, B. K., & Newell, B. R. (2011). Noncategorical approaches to feature prediction with uncertain categories. *Memory & Cognition*, *39*, 304–318. doi:10.3758/s13421-010-0009-4
- Patalano, A. L., & Ross, B. H. (2007). The role of category coherence in experience-based prediction. *Psychonomic Bulletin & Review*, *14*, 629–634. doi:10.3758/BF03196812
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308. doi:10.1037/h0028558
- Rehder, B. (2006). When causality and similarity compete in category-based property induction. *Memory & Cognition*, *34*, 3–16. doi:10.3758/BF03193382
- Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition*, *91*, 113–153. doi:10.1016/S0010-0277(03)00167-7
- Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681. doi:10.1016/S0022-5371(75)80055-7
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–656.
- Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In K. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning*. Cambridge, United Kingdom: Cambridge University Press.
- Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189–228. doi:10.1207/s15516709cog2202\_2
- Verde, M. F., Murphy, G. L., & Ross, B. H. (2005). Influence of multiple categories on the prediction of unknown properties. *Memory & Cognition*, *33*, 479–487. doi:10.3758/BF03193065
- Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 776–795. doi:10.1037/0278-7393.26.3.776

(Appendices follow)

## Appendix A

The entropy of each stimulus dimension (e.g.,  $X$ ) was calculated by supposing dimension  $X$  has  $n$  values (e.g.,  $x_1, x_2 \dots x_n$ ), each of which can occur with a particular probability,  $p(x_i)$ . The entropy in dimension  $X$  (in bits) is then calculated as follows:

$$H(X) = \sum_{i=1}^n -p(x_i) \log_2 p(x_i) \dots \quad (\text{A1})$$

The minimum value is always 0, but the maximum value depends on the number of values per dimension ( $n$ ) and, therefore, differs across stimulus sets. The maximum value is given by the following:

$$H(X)_{\max} = \log_2 n.$$

Mutual information refers to the reduction in entropy in dimension  $X$  conditional on dimension  $Y$ . We separately calculated the

reduction in entropy of each stimulus dimension ( $X$ ) conditional on the category label dimension,  $Cat$ , which can take values  $[y_1, y_2 \dots y_m]$ :

$$I(X, Cat) = H(X) - H(X|Cat) \dots \quad (\text{A2})$$

where

$$H(X|Cat) = \sum_{j=1}^m \sum_{i=1}^n -p(x_i|y_j) \log_2 p(x_i|y_j) \dots \quad (\text{A3})$$

These information values were then averaged over the number of stimulus dimensions to yield a mean information value of the category labels with respect to all stimulus feature dimensions. The resultant values are reported in Table A1.

Table A1  
*Information Values of Stimuli in Prior Experiments*

Study	$H(X)$ (bits)	$I(Cat, X)$ (bits)	Exemplars repeated?	Stimulus dimensions
Present study				
High coherence	2.1361	0.9385	No	$5 \times 5 \times 5 \times 5 \times 5$
Low coherence	2.2741	0.3288	No	$5 \times 5 \times 5 \times 5 \times 5$
Papadopoulos et al. (2011)				
Experiment 1	1.528213	0.109759	Yes	$3 \times 3$
Experiment 2	1.480343	0.037447	Yes	$3 \times 3$
Experiment 3	1.528213	0.109759	Yes	$2 \times 3 \times 3$
Experiment 4				
2 categories	1.680482	0.062028	Yes	$3 \times 4$
3 categories	1.685241	0.119981	Yes	$3 \times 4$
Murphy & Ross (2010a)	1.824088	0.471871	Yes	$7 \times 9$
Murphy & Ross (2010b)				
50–50	1.963698	0.96587	Yes	$6 \times 6$
60–40	1.990202	0.881743	Yes	$6 \times 6$
80–20	1.995538	1.055418	Yes	$5 \times 6$
Murphy & Ross (2005)				
Cue validity	1.991646	0.922027	Yes	$4 \times 4 \times 4$
Category validity	1.982039	0.897077	Yes	$4 \times 4 \times 4$
Murphy & Ross (1994)				
Experiment 1	1.766934	0.869566	Yes	$3 \times 4$
Experiment 4	1.579434	0.736295	Yes	$3 \times 3$
Experiments 5 and 6	1.781601	0.535486	Yes	$3 \times 4$
Experiment 8	1.939114	0.658475	Yes	$4 \times 4$
Experiment 10	1.880241	1.099602	Yes	$4 \times 4$
Experiment 11	1.766934	0.611295	Yes	$3 \times 4$
Verde et al. (2005)				
Experiments 1 and 2	1.51897	0.621602	Yes	$3 \times 4$
Experiment 3	2.163534	0.792583	Yes	$5 \times 6$
Newell et al. (2010)	2.214793	0.378478	Yes	$5 \times 6$
Hayes & Newell (2009)				
Divergent	1.71059	0.64859	Yes	$4 \times 4$
Nondivergent	1.414717	0.360488	Yes	$3 \times 3$
Griffiths et al. (in press)	1.796285	0.717029	Yes	$4 \times 4$

*Note.* Summary of category coherence and number/size of dimensions used in existing uncertain induction studies.  $H(X)$  refers to the mean entropy of the stimulus set and can be considered a measure of stimulus complexity.  $I(Cat, X)$  refers to the mean information provided by a category label about the feature dimensions of the exemplars in that category. It can be considered a measure of category coherence. Note that these information values are artificially high for stimulus sets in which identical exemplars are repeated (not the present stimuli). Both entropy and mutual information are measured in bits, with higher numbers indicating more entropy or mutual information.

(Appendices continue)

## Appendix B

## Category Structures Used in Experiments 1 and 2

Coherence	Category	Exemplar no.	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
High	Target	1	1	4	1	1	1
High	Target	2	1	4	2	1	1
High	Target	3	1	4	4	1	1
High	Target	4	1	4	5	1	1
High	Target	5	1	4	1	2	1
High	Target	6	1	5	1	1	1
High	Target	7	1	5	1	4	1
High	Target	8	1	1	1	5	1
High	Target	9	1	1	1	1	2
High	Target	10	1	1	1	1	4
High	Target	11	3	1	1	1	1
High	Target	12	3	1	1	1	5
High	Target	13	4	1	1	1	1
High	Target	14	4	1	4	1	1
High	Target	15	5	1	1	1	1
High	Target	16	5	1	1	5	1
High	Nontarget	1	1	5	2	2	2
High	Nontarget	2	1	5	4	2	2
High	Nontarget	3	1	5	3	2	2
High	Nontarget	4	1	5	2	4	2
High	Nontarget	5	1	5	2	3	2
High	Nontarget	6	2	2	1	2	2
High	Nontarget	7	2	2	5	2	2
High	Nontarget	8	2	2	2	1	2
High	Nontarget	9	2	2	2	5	2
High	Nontarget	10	2	2	2	2	1
High	Nontarget	11	2	2	2	2	3
High	Nontarget	12	2	2	2	2	4
High	Nontarget	13	2	2	2	2	5
High	Nontarget	14	2	2	4	2	2
High	Nontarget	15	2	2	2	4	2
High	Nontarget	16	5	2	2	2	2
High	Irrelevant	1	4	3	3	3	3
High	Irrelevant	2	5	3	3	3	3
High	Irrelevant	3	3	4	3	3	3
High	Irrelevant	4	3	5	3	3	3
High	Irrelevant	5	3	3	4	3	3
High	Irrelevant	6	3	3	5	3	3
High	Irrelevant	7	3	3	3	4	3
High	Irrelevant	8	3	3	3	5	3
High	Irrelevant	9	3	3	3	3	4
High	Irrelevant	10	3	3	3	3	5
High	Irrelevant	11	2	3	3	3	3
High	Irrelevant	12	3	1	3	3	3
High	Irrelevant	13	3	3	2	3	3
High	Irrelevant	14	3	3	3	1	3
High	Irrelevant	15	3	3	3	3	2
High	Irrelevant	16	3	3	3	2	3
Low	Target	1	1	4	2	1	5
Low	Target	2	1	4	2	1	3
Low	Target	3	1	4	5	2	1
Low	Target	4	1	4	4	3	1
Low	Target	5	1	4	1	4	3
Low	Target	6	1	5	1	4	1
Low	Target	7	1	5	4	1	2

(Appendices continue)

Appendix B (*continued*)

Coherence	Category	Exemplar no.	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
Low	Target	8	1	1	4	5	1
Low	Target	9	1	1	5	5	2
Low	Target	10	1	1	5	1	2
Low	Target	11	4	1	1	1	3
Low	Target	12	4	1	1	1	4
Low	Target	13	5	1	4	2	1
Low	Target	14	5	1	2	3	1
Low	Target	15	2	1	1	4	5
Low	Target	16	3	1	1	5	4
Low	Nontarget	1	1	5	2	4	2
Low	Nontarget	2	1	5	2	3	2
Low	Nontarget	3	1	5	4	2	2
Low	Nontarget	4	1	5	2	2	3
Low	Nontarget	5	1	5	2	2	2
Low	Nontarget	6	2	2	2	4	1
Low	Nontarget	7	2	2	1	5	2
Low	Nontarget	8	2	2	1	5	3
Low	Nontarget	9	2	3	4	2	5
Low	Nontarget	10	2	3	4	2	1
Low	Nontarget	11	2	4	3	2	5
Low	Nontarget	12	2	2	3	5	4
Low	Nontarget	13	3	1	2	4	2
Low	Nontarget	14	4	2	5	2	1
Low	Nontarget	15	4	2	5	2	3
Low	Nontarget	16	5	2	3	2	4
Low	Irrelevant	1	3	3	1	2	4
Low	Irrelevant	2	3	1	3	2	4
Low	Irrelevant	3	3	2	4	3	5
Low	Irrelevant	4	3	2	4	5	3
Low	Irrelevant	5	4	3	3	5	1
Low	Irrelevant	6	5	3	1	3	2
Low	Irrelevant	7	4	3	5	1	3
Low	Irrelevant	8	2	1	3	3	4
Low	Irrelevant	9	2	4	3	5	3
Low	Irrelevant	10	5	1	2	3	3
Low	Irrelevant	11	3	4	5	3	1
Low	Irrelevant	12	3	3	5	1	3
Low	Irrelevant	13	2	3	3	4	5
Low	Irrelevant	14	5	1	3	3	2
Low	Irrelevant	15	2	1	4	3	3
Low	Irrelevant	16	3	4	5	1	3

*Note.* Each number 1 through 5 indicates a different feature value. The given feature for each uncertain induction problem was Feature Value 1 on Dimension 1. The feature participants needed to infer was the value of Dimension 2. Dimensions 3 through 5 were not directly involved in feature inference. Dim = Dimension.

Received January 16, 2011  
Revision received September 1, 2011  
Accepted September 8, 2011 ■