COMMENTARY

An Exemplar-Model Account of Feature Inference From Uncertain Categorizations

Robert M. Nosofsky
Indiana University Bloomington

In a highly systematic literature, researchers have investigated the manner in which people make feature inferences in paradigms involving uncertain categorizations (e.g., Griffiths, Hayes, & Newell, 2012; Murphy & Ross, 1994, 2007, 2010a). Although researchers have discussed the implications of the results for models of categorization and inference, an explicit formal model that accounts for the full gamut of results has not been evaluated. Building on previous proposals, in this theoretical note I consider the predictions from an exemplar model of categorization in which the inferred category label becomes a new feature of the objects. The model predicts a priori a wide range of robust results that have been documented in this literature and can also be used to interpret effects of experimental manipulations that modulate these results. The model appears to be an excellent candidate for understanding the manner in which specific exemplar information and category inferences are combined to generate inferences about new features of objects.

Categorization is a fundamental cognitive process that brings organization and efficiency to people's mental lives. One of its major functions is that it facilitates the making of inferences. If one learns that an object is a member of the category "fruit," then a reasonable inference is that the object has seeds. However, inferences about unknown properties are also made on the basis of specific exemplar-level knowledge. For example, observation of an object with a specific color and texture may lead the observer to judge that it is a watermelon of the seedless variety.

A central question in cognitive psychology and cognitive science concerns how people combine information about specific exemplars and judged category membership when making inferences. In highly systematic programs of research, Murphy, Ross, and their colleagues (Murphy, Chen, & Ross, 2012; Murphy and Ross, 1994, 2007, 2010a, 2010b) and Hayes, Newell, and their colleagues (Griffiths et al., 2012; Hayes & Newell, 2009; Newell, Paton, Hayes, & Griffiths, 2010; Papadopoulos, Hayes, & Newell, 2011) have pursued this question under conditions in which category-membership information is uncertain. The paradigms used in these studies are variants and extensions of the one introduced by Murphy and Ross (1994) and illustrated in Figure 1. In the basic paradigm, subjects view a set of specific exemplars that are organized into categories. In the Figure 1 example, the exemplars are colored shapes, and the categories are the children (Bob, John, Sam, and Ed) who drew these shapes. The experimenter informs the subject that a child has drawn a new picture with a particular shape (say, triangle); the subject is then asked to predict which child most likely drew the picture as well as the color of the new picture.

Extensive research with this paradigm has revealed several major trends in performance. First, subjects often (but not always) seem to place emphasis on a single inferred category in making their inferences. In the Figure 1 example, if told that a child has drawn a new triangle, the most likely inference is that the child was Bob (because Bob tends to draw triangles). And if the subject makes this inference, then he or she will typically judge that the color of the new object is likely to be black (because Bob tends to draw black figures). Furthermore, this color inference seems often (but not always) to be unaffected by the extent to which other children draw black figures. Instead, the focus in reaching the color inference is on the single category (Bob).

A second major trend is that subjects are also sensitive to feature conjunctions. For example, in Figure 1, not only does Bob tend to draw black figures, but the majority of Bob's *triangles* are black.

In the Figure 1 example, the inferred color would tend to be black regardless of whether the subject made an independent count of the total number of objects that were black or made a count that was conditional on the shape being a triangle (i.e., a feature-conjunction strategy). As will be seen, however, other category structures have been designed that do make this contrast, and the evidence points to a highly significant role of feature conjunctions in making inferences. As discussed and demonstrated by Griffiths et al. (2012), in some cases the strategy seems to involve a *single-category feature-conjunction* strategy (e.g., count the number of black triangles only within the Bob category), whereas in others it seems to involve a *multiple-category feature-conjunction* strategy (count the number of black triangles across all categories, not just the inferred one).

In discussing the implications of these patterns of results, Murphy and Ross (1994, 2007, 2010a) noted that they challenge the predictions of a variety of formal models of categorization and inference. However, to my knowledge, they have never proposed and tested a particular formal model that *does* capture their com-

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405. E-mail: nosofsky@indiana.edu
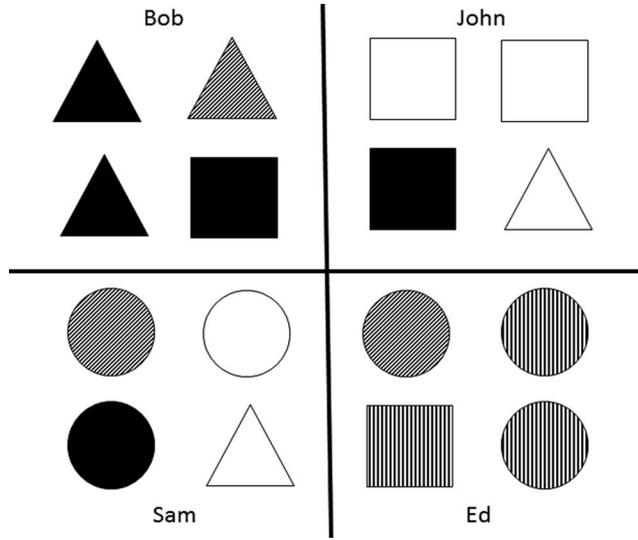
Figure 1. Design from Experiment 1 of Murphy and Ross (1994). Adapted from "Predictions from uncertain categorizations," by G. L. Murphy and B. H. Ross, 1994, *Cognitive Psychology, 27,* pp. 148–193, Copyright 1994 by Elsevier. Adapted with permission.

plete gamut of results. Griffiths et al. (2012) provided an enlightening discussion to suggest that certain types of exemplar models seem consonant with the results they reported in their 2012 article. However, they did not explicitly formalize such a model to derive its predictions and did not consider the predictions of such a model for the many earlier studies of Murphy and Ross.

The purpose of this theoretical note is to fill this gap and propose an explicit formal model for inference in these tasks. In my view, this direction is an important one, because the very goal of science is to provide explanations for phenomena of interest. Murphy and Ross have made highly significant contributions in pruning the "search tree" of explanations by noting how their results rule out various models. Here, I take a complementary tack by following what appears to be a highly promising branch of the tree.

The proposed model is a straightforward extension of a standard exemplar model (Medin & Schaffer, 1978; Nosofsky, 1984) of the form sketched by Griffiths et al. (2012, pp. 590–591). The extended model has its roots in earlier proposals made, for example, by Anderson (1990), Heit (1992), Nosofsky and Johansen (2000, p. 397), and Sloutsky and Fisher (2004). The key idea is that when an observer makes an inference that an object belongs to a category, the inferred category label becomes a new feature of that object. The observer then assesses the similarity of the object/category-label *pair* to all other object-label pairs in the display, using the similarity rules formalized in the exemplar models of Medin and Schaffer (1978) and Nosofsky (1984). The probability that the subject infers that the newly queried feature is Feature X is then based on the summed similarity of the object-label pair to all object-label exemplars that contain Feature X (Medin & Schaffer, 1978; Nosofsky, 1984).

The Formal Model and an Application to the Basic Phenomenon

To illustrate, consider again the Figure 1 example. In their Experiment 1, Murphy and Ross (1994) contrasted probability

judgments for two types of inferences. In their *increasing condition,* the given feature was triangle and the main dependent variable was subjects' probability judgment for the feature "black." Following Murphy and Ross (1994), I refer to the category that is the most likely inference as the *target category* (in this example, Bob). Note that three of the four figures in the target category are black. This condition is referred to as *increasing* because other plausible categories besides the target category also provide some evidence that the color of the new drawing may be black: Sam and John also sometimes draw triangles, and these children sometimes also make drawings that are black. By contrast, in the *neutral condition,* the given feature was square, and the dependent variable was subjects' probability judgment for the feature "white." In this case, the target category is John, and analogous to the increasing condition, three of the four figures in the target category point to white. However, in this neutral condition, there is no additional evidence from plausible alternative categories that the color is likely to be white: Bob and Ed also sometimes draw squares, but their drawings are never white. Thus, Murphy and Ross reasoned that to the extent that subjects make use of information from multiple categories, the probability judgments for the critical feature should be greater in the increasing condition than in the neutral condition. Their central finding, however, which they replicated in several experiments involving this structural design, was that there was no difference in the probability judgments for the critical feature across the increasing and neutral conditions. Instead, it appeared that the feature inferences were with respect to only the single target category.

To apply the proposed exemplar model to this design, I assume that subjects did indeed infer the target category when presented with the given feature. (Murphy and Ross generally restrict their analyses to that vast majority of subjects who do indeed infer the target category when asked the category question.) Thus, using the notation scheme in Table 1, in the increasing condition, the object-label pair that serves as the probe is 1?1, where 1 on Dimension 1 is the value of the given shape feature (triangle); 1 on Dimension 3 is the inferred label (Bob); and "?" on Dimension 2 denotes the missing color feature that is being queried. Following Medin and Schaffer (1978), the probability that the subject judges that the color feature is black is then found by summing the similarity (s) of the 1?1 probe to all object-label pairs that are black, and dividing by the summed similarity of the 1?1 pair to all object-label pairs irrespective of the color feature. (Note that in the proposed model, these summed similarities are computed across all four children categories, not just the inferred one.) Denoting black as Value 1 on Dimension 2 (see Table 1), I denote the set of all such black exemplars as $\mathbf{j} \, \varepsilon \, D_2(1)$. Thus, according to the model,

Table 1
Feature Notation for Illustrating the Application of the Exemplar Model to the Murphy and Ross (1994) Paradigm

Shape	Color	Category label
1 = triangle	1 = black	1 = Bob
2 = square	2 = diagonal stripe	2 = John
3 = circle	3 = white (or empty)	3 = Sam
	4 = vertical stripe	4 = Ed

$$P(\text{black}|\text{triangle}) = \sum_{\mathbf{j} \varepsilon D_2(1)} s[1 \, ? \, 1, \mathbf{j}] / \sum_{\mathbf{j}} s[1 \, ? \, 1, \mathbf{j}], \qquad (1)$$

where the denominator in Equation 1 denotes the summed similarity of object-label pair 1?1 to all exemplars of all categories.

Similarity is computed using the well-known interdimensional-multiplicative rule from the exemplar model (Medin & Schaffer, 1978; Nosofsky, 1984).[1] For example, the similarity of probe x?y to exemplar u?v is given by

$$s(x \, ? \, y, u \, ? \, v) = s(x, u) * s(y, v), \qquad (2)$$

where the similarity along Dimension 1 (the shape dimension) is given by $s(x, u) = S$ ($0 \leq S \leq 1$) if the items mismatch on Dimension 1, and $s(x, u) = 1$ if they match. Likewise, the similarity along Dimension 3 (the category label dimension) is given by $s(y, v) = L$ ($0 \leq L \leq 1$) if the items mismatch on the label dimension, and $s(y, v) = 1$ if they match. (Recall that more salient feature mismatches are represented by *smaller* values of the similarity parameters.)

Thus, in the increasing condition described above, P(black|triangle) is given by

$$P(\text{black}|\text{triangle}) = [2 + S + 2SL]/[(3 + S) + (L + 3SL)$$
$$+ (L + 3SL) + (4SL)], \qquad (3a)$$

where, for clarity, the terms in the denominator are grouped according to category (Bob, John, Sam, and Ed in Figure 1). For example, the pair triangle_Bob (1?1) has similarity 1 to the two black triangles in the Bob category, similarity S to the black square in the Bob category, and similarity SL to the black square in the John category and the black circle in the Sam category. Thus, the summed similarity to all black figures (the numerator in Equation 3a) is $2 + S + 2SL$.

For the neutral condition, the given feature (square) leads the observer to infer that the target category is John; thus, using the Table 1 notation, the relevant probe pair is 2?2. To predict the probability that the inferred feature is white, one sums the similarity of the 2?2 probe to all white exemplars in Figure 1, and divides by the summed similarity of the 2?2 probe to all exemplars. This computation yields

$$P(\text{white}|\text{square}) = [2 + S + 2SL]/[(3 + S) + (L + 3SL)$$
$$+ (L + 3SL) + (4SL)]. \qquad (3b)$$

Comparison of the expressions in Equations 3a and 3b reveals that these probability computations are identical. Thus, the proposed model predicts a priori Murphy and Ross's robust finding that the probability judgments across the increasing and neutral conditions were essentially identical. It is important to note, however, that the present formal account does not posit that observers give no consideration to members of alternative categories, as is true in the account suggested by Murphy and Ross (1994) for their early experiments. Instead, the extent to which evidence from other categories plays a role would be influenced by the magnitude of the label-mismatch parameter L. If the label dimension is given a great deal of "attention weight" (Nosofsky, 1984), so that L is close to zero, then evidence from other categories would indeed be "zeroed out." However, if observers give less weight to the in-

ferred label dimension, then evidence from alternative categories could play a role. For the present design, it turns out that regardless of the magnitude of L, the probability predictions for the inferred color are identical across the increasing and neutral conditions. As will be seen, however, the model is also able to account for performance in designs that indicate that evidence from alternative categories does sometimes play a role.

Simultaneous Manipulation of Target-Category Base Rates and Other-Category Evidence

The next major design structures considered by Murphy and Ross (1994) were the ones tested in their Experiment 4, two of which are illustrated here as Figure 2. Murphy and Ross manipulated two variables in this experiment. The first variable was the "increasing" versus "neutral" variable that was manipulated in Experiment 1, that is, whether exemplars from alternative plausible categories did or did not provide additional evidence for the critical to-be-predicted feature. The second variable was the strength of evidence from the target category itself, that is, the number of exemplars in the target category that have the critical to-be-predicted feature. In a *high base-rate condition*, three exemplars from the target category had the critical to-be-predicted feature, whereas in the *low base-rate condition*, only two exemplars had this critical feature. Murphy and Ross's purpose in conducting the experiment was to test their prediction that the null effect of the increasing/neutral variable would be observed in combination with a significant effect of the base-rate manipulation.

An example of the high-base-rate/neutral condition is illustrated in Figure 2 for the case in which the given feature is triangle. The inferred category is Bob, and three of the four figures in Bob's category are black (high base rate). Jim and Sam also sometimes draw triangles, but they never draw black figures, so evidence from these plausible alternative categories does not increase the inference that the color may be black (neutral condition). The low-base-rate/neutral condition was identical to the one just described, except one of the black triangles in the target category would be switched to a triangle of a different color.

An example of the low-base-rate/increasing condition is shown in Figure 2 for the case in which the given feature is circle. The inferred target category is Sam, and two of the exemplars in the Sam category are white (low base rate). Jim and Bob also sometimes draw circles, and some of their figures are white, which might increase the probability estimate for the to-be-predicted feature (increasing condition). The high-base-rate/increasing condition was identical to the one just described, except one of Sam's shaded circles would be switched to one that was white, so three exemplars in the target category would be white.

As predicted by Murphy and Ross, the base-rate manipulation had a big effect on the feature inferences, with subjects predicting the critical feature far more often in the high-base-rate conditions than the low-base-rate ones. By contrast, regardless of base rate,

---

[1] Because the present designs are limited to cases involving stimuli varying along discrete dimensions, the simple multiplicative-similarity rule from Medin and Schaffer (1978) suffices and Nosofsky's (1984) exponential-similarity interpretation is not needed.
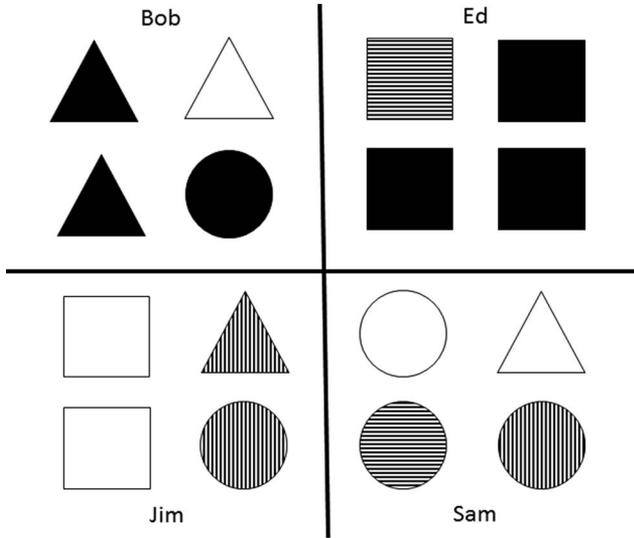
Figure 2. Design from Experiment 4 of Murphy and Ross (1994). Adapted from "Predictions from uncertain categorizations," by G. L. Murphy and B. H. Ross, 1994, *Cognitive Psychology, 27,* pp. 148–193, Copyright 1994 by Elsevier. Adapted with permission.

the increasing/neutral manipulation had no effect on the inferences, replicating the main result from Experiment 1.

Following the computational machinery described in detail for Experiment 1, the derived exemplar-model prediction equations across the four conditions are as follows (the subscript on P denotes the base-rate condition):

High-base-rate/neutral

$$P_H(\text{black}\,|\,\text{triangle}) = [2 + S + 3SL]/[(3 + S) + (4SL)$$
$$+ (L + 3SL) + (L + 3SL)]. \quad (4a)$$

Low-base-rate/neutral

$$P_L(\text{black}\,|\,\text{triangle}) = [1 + S + 3SL]/[(3 + S) + (4SL)$$
$$+ (L + 3SL) + (L + 3SL)]. \quad (4b)$$

High-base-rate/increasing

$$P_H(\text{white}\,|\,\text{circle}) = [2 + S + 3SL]/[(L + 3SL) + (4SL)$$
$$+ (L + 3SL) + (3 + S)]. \quad (5a)$$

Low-base-rate/increasing

$$P_L(\text{white}\,|\,\text{circle}) = [1 + S + 3SL]/[(L + 3SL) + (4SL)$$
$$+ (L + 3SL) + (3 + S)]. \quad (5b)$$

In these equations, to increase clarity, the terms in the denominators are grouped in order of the illustrated categories (Bob, Ed, Jim, Sam).

Comparison of Equation 4a with Equation 4b and Equation 5a with Equation 5b reveals that, regardless of the level of the increasing/neutral manipulation and regardless of parameter settings, the exemplar model predicts that the critical feature infer-

ences will be greater in the high-base-rate conditions than the low-base-rate ones, in agreement with Murphy and Ross's findings. At the same time, comparison of Equation 4a with Equation 5a and Equation 4b with Equation 5b reveals that the feature-inference predictions are identical across the increasing and neutral conditions, again in agreement with Murphy and Ross's findings. In sum, the proposed model predicts simultaneously the observed effect of the base-rate manipulation and the null effect of the increasing/neutral manipulation in this design.

Pitting the Single-Category View Against a Particular Multiple-Category Bayesian Model

The previous designs relied on a finding of "no difference" between the increasing and neutral conditions to suggest that subjects focused on the target category in making their critical-feature inferences. In a subsequent design, Murphy and Ross (1994, Experiments 5 and 6) created a new structure that yielded a sharp qualitative contrast between the predictions from the single-category view and a particular multiple-category Bayesian model of feature inference (see Murphy & Ross, 1994, pp. 166–167 for details). The structure is illustrated here as Figure 3. An example of the key comparison arises when the given feature is triangle, in which case the target category is Bob. Given this target-category inference, the single-category view predicts that the inferred color will be black, because 50% of the objects in the Bob category are black, and only 33% of the objects in the Bob category are white (with remaining colors being even less frequent). By contrast, the particular multiple-category Bayesian model considered by Murphy and Ross predicted that the inferred feature would be white. Finally, an alternative pure "feature-conjunction" strategy was also considered. According to this strategy, subjects simply count the total number of triangles that are black and the total number of triangles that are white, without respect to likely category membership. Note that in the Figure 3 design, there are equal numbers of black and white triangles. Thus, this model predicted no difference in inferences for black versus white. As it turned out, the general pattern of results observed by Murphy and Ross (1994) favored the predictions from the single-category view, with most subjects inferring the feature black. This pattern was exceptionally strong in the version of the experiment in which subjects first identified the likely target category.

When applied to the Figure 3 design, the prediction equations yielded by the exemplar model are as follows:

"*single-category inference*"

$$P(\text{black}\,|\,\text{triangle}) = [3 + 4SL]/[(4 + 2S) + (6SL) + (6SL)$$
$$+ (2L + 4SL)]; \quad (6a)$$

"*multiple-category Bayesian inference*"

$$P(\text{white}\,|\,\text{triangle}) = [1 + S + 2L + 3SL]/[(4 + 2S) + (6SL)$$
$$+ (6SL) + (2L + 4SL)]. \quad (6b)$$

The denominators for Equations 6a and 6b are identical. Furthermore, the reader can verify that across its entire parameter space (excluding the degenerate case $L = 1$), the exemplar-model numerator is greater in Equation 6a than in Equation 6b, so it
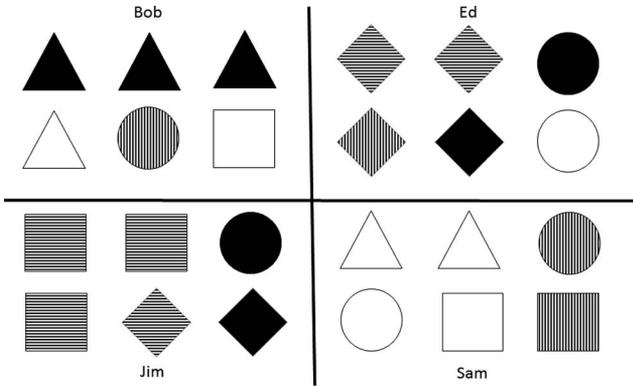
Figure 3. Design from Experiments 5–6 of Murphy and Ross (1994). Adapted from "Predictions from uncertain categorizations," by G. L. Murphy and B. H. Ross, 1994, *Cognitive Psychology, 27,* pp. 148–193, Copyright 1994 by Elsevier. Adapted with permission.

predicts P(black | triangle) > P(white | triangle). Thus, the model predicts a priori the qualitative pattern of results observed by Murphy and Ross for this design.

The Role of Feature Correlations

In pursuing further the nature of subjects' feature inferences, Murphy and Ross (1994) conducted additional experiments that examined the role of feature correlations. Their first critical design for examining this issue was their Experiment 8, depicted here as Figure 4. In one condition, the given feature was circle, and at issue was the extent to which subjects would predict that the likely value on the color dimension was "vertically striped." In a comparison condition, the given feature was triangle, and at issue was the extent to which subjects would predict that the value on the color dimension was
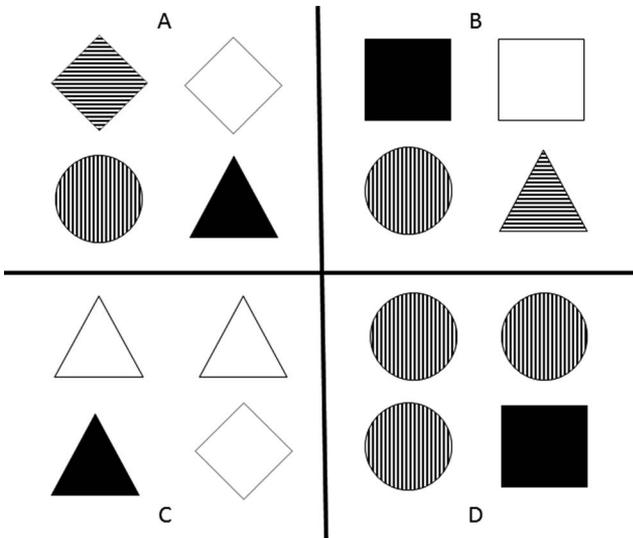

Figure 4. Design from Experiment 8 of Murphy and Ross (1994). Adapted from "Predictions from uncertain categorizations," by G. L. Murphy and B. H. Ross, 1994, *Cognitive Psychology, 27,* pp. 148–193, Copyright 1994 by Elsevier. Adapted with permission.

white. Murphy and Ross argued that a model that was sensitive to correlated features would predict P(vertical stripe | circle) > P(white | triangle). The reason is that all circles in the Figure 4 design are vertically striped, whereas only a subset of triangles are white, so the correlation between circle and vertically striped is much stronger than between triangle and white. By contrast, the particular independent-feature Bayesian model considered by Murphy and Ross predicted no difference in these feature inferences (see Murphy & Ross (1994), p. 173). Murphy and Ross's results strongly supported the predictions from the correlated-features view, with the probability judgments for P(vertical-stripe | circle) far exceeding the probability judgments for P(white | triangle).

Again, it is straightforward to apply the proposed exemplar model to this design. Note that in the case in which the given feature is circle, the inferred category is D; whereas in the case in which the given feature is triangle, the inferred category is C. Thus, in the first case the feature-label pair is circle_D whereas in the second it is triangle_C. Given these assumptions, the prediction equations yielded by the exemplar model are:

Strong correlation

$$P(\text{vertical} - \text{stripe}\,|\,\text{circle}) = [3 + 2L]/[(L + 3SL) + (L + 3SL)$$

$$+ (4SL) + (3 + S)]. \qquad (7a)$$

Weak correlation

$$P(\text{white}\,|\,\text{triangle}) = [2 + S + 2SL]/[(L + 3SL) + (L + 3SL)$$

$$+ (3 + S) + (4SL)]. \qquad (7b)$$

The denominators of Equations 7a and 7b are identical; and across the model's parameter space, the numerator of Equation 7a is greater than the numerator of Equation 7b. Thus, the exemplar model predicts a priori Murphy and Ross's result that P(vertical-stripe | circle) > P(white | triangle), indicating that the model is indeed appropriately sensitive to correlated features in this feature-inference design.

Potential Effects of a High-Plausibility Single Alternative Category

In further experiments, Murphy and Ross (1994) considered potential limitations of their hypothesis that subjects focus on a single target category in making inferences to new features. In their Experiment 10, they created conditions in which a single alternative category (besides the target one) was or was not highly likely. The intuition is that if a single alternative category is highly likely, then subjects might integrate information from both categories rather than focusing solely on the target one.

The design they used to test this hypothesis is shown in Figure 5. In the condition in which a single alternative category was highly likely, the given feature was triangle and the question was the extent to which subjects would judge that the likely color was black. In this case, the target category is Bob, and three of Bob's triangles are black. However, the alternative category Jim is also highly likely, but none of Jim's drawings are black. Thus, the intuition is that to the extent that subjects are also pulled toward the alternative category, the judgments for the feature black might be weakened. In a comparison condition, the given feature was white and the question was the extent to which subjects would judge that the
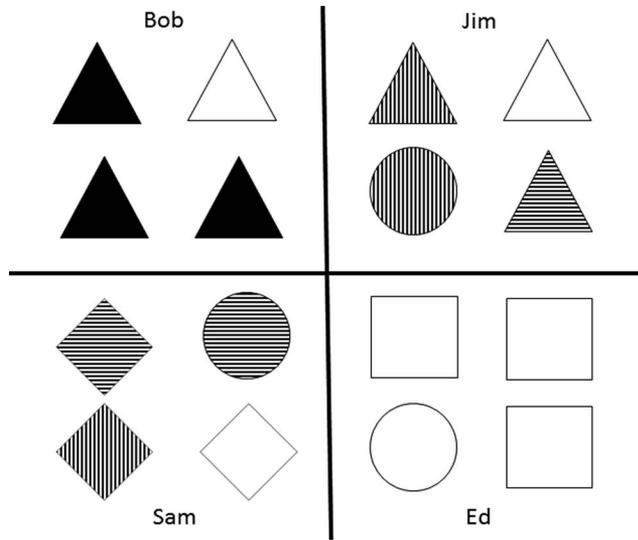
Figure 5. Design from Experiment 10 of Murphy and Ross (1994). Adapted from "Predictions from uncertain categorizations," by G. L. Murphy and B. H. Ross, 1994, *Cognitive Psychology, 27,* pp. 148–193, Copyright 1994 by Elsevier. Adapted with permission.

likely shape is square. In this case, the target category is Ed, and three of Ed's figures are squares. However, there is now no single alternative category that is highly likely, so the judgment for square would not be weakened. Thus, to the extent that subjects are sometimes drawn to a single highly likely alternative category, the hypothesis is that P(black | triangle) < P(square | white). However, Murphy and Ross (1994) observed no difference in the judgments across these two conditions, lending still further support to the single-category hypothesis of feature inference.

For the Figure 5 design, the exemplar-model predictions are as follows:

Single high-likelihood alternative

$$P(\text{black} \mid \text{triangle}) = [3]/[((4) + (3L + SL) + (4SL) + (4SL)]$$
$$= [3]/[4 + 3L + 9SL].$$

(8a)

No single high-likelihood alternative

$$P(\text{square} \mid \text{white}) = \frac{[3]/[(L + 3CL) + (L + 3CL)}{+ (L + 3CL) + (4)]}$$
$$= [3]/[4 + 3L + 9CL].$$

(8b)

Note that in Equation 8b, the parameter C denotes the similarity between mismatching values on the color dimension rather than the shape dimension. Assuming $S = C$ (i.e., that the color and shape dimensions are equally salient), then the expressions in Equations 8a and 8b are identical. Furthermore, Murphy and Ross balanced across their conditions which feature dimension played each logical role in the abstract design. Thus, the exemplar model again predicts a priori Murphy and Ross's major qualitative result: The critical feature inferences are predicted to be identical across the conditions in which a single alternative category is or is not highly likely.

The Role of Feature Conjunctions Outside the Target Category

In the final experiment of their 1994 study (Experiment 11), Murphy and Ross made another attempt to find evidence of use of multiple categories beyond the target one. In this case, the idea was to use correlated feature attributes that occurred *outside* the target category. (In their earlier Experiment 8, the correlated features occurred within the target category.) The design of the experiment is illustrated in Figure 6. In the *conjunction* condition, the given feature was triangle. The target category is Arnie (who drew three triangles), so the critical to-be-predicted feature is black (three of Arnie's drawings are black). Neil and Roger also sometimes draw triangles, so these are alternative plausible categories. Crucially, in these alternative categories, the triangles are also black. Thus, in this condition, the critical feature conjunctions occur outside the target category. In a comparison *independent* condition, the probe feature was white. The likely target category is Gary (three white drawings), so the likely to-be-predicted feature is square (Gary drew three squares). Neil and Roger also drew white figures, so they are plausible alternative categories. Furthermore, these children also drew squares. Crucially, however, in these alternative categories, the to-be-predicted feature (square) occurs in separate exemplars from the given feature (white), breaking the feature conjunction. As described by Murphy and Ross (1994, pp. 181–182), both the single-category view as well as their independent-feature Bayesian model predict no difference in feature inferences and probability judgments across the correlated and independent conditions. In contrast to these predictions, inferences and probability judgments for the critical features were significantly greater in the conjunction condition than in the independent condition. Murphy and Ross (1994, p. 182) characterized these results as surprising for a wide range of reasons and stated that ". . . it is quite clear that the particular
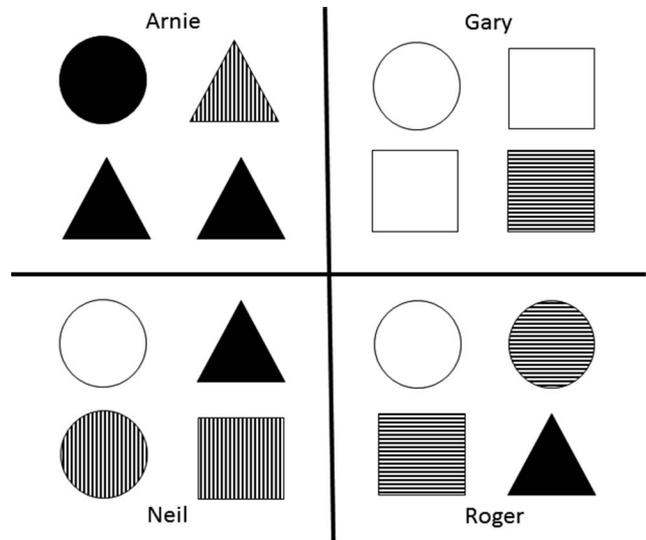

Figure 6. Design from Experiment 11 of Murphy and Ross (1994). Adapted from "Predictions from uncertain categorizations," by G. L. Murphy and B. H. Ross, 1994, *Cognitive Psychology, 27,* pp. 148–193, Copyright 1994 by Elsevier. Adapted with permission.

pairing of features outside the target category is having an important effect."

The prediction equations from the exemplar model for these two conditions are as follows:

Conjunction

$$P(\text{black}|\text{triangle}) = [2 + S + 2L]/[(3 + S) + (4SL) + (L + 3SL)$$

$$+ (L + 3SL)]. \quad (9a)$$

Independent

$$P(\text{square}|\text{white}) = [2 + C + 2CL]/[(4CL) + (3 + C)$$
$$+ (L + 3CL) + (L + 3CL)]$$
$$= [2 + C + 2CL]/[(3 + C) + (4CL)$$
$$+ (L + 3CL) + (L + 3CL)]$$

$$(9b)$$

Assuming equal feature salience ($C = S$), the denominators of Equations 9a and 9b are identical, and the numerator of Equation 9a always exceeds the numerator of Equation 9b (because $2L > 2CL$). Thus, the exemplar model predicts a priori Murphy and Ross's result that inferences to the critical feature were greater in the conjunction condition than in the independent condition.

More Evidence for a Role of Feature Conjunctions

In more recent work, Murphy and Ross (2010a) pursued further the role that feature conjunctions may play in inference. They conducted a design that pitted feature conjunctions within a target category against category-level information based on independent features. The design is illustrated in Figure 7. The crucial results arise in what Murphy and Ross (2010a) refer to as their *disagree* condition. In an example of this condition, the given feature would be yellow. The target category is Maura, who drew five yellow figures. At issue was the extent to which subjects would infer that the shape of the new drawing was likely to be a circle versus a diamond. If subjects focus mainly on the yellow figures within the Maura category (a *feature-conjunction* strategy), then the likely inference is that the shape will be a circle (most of the yellow figures in the Maura category are circles). But if subjects evaluate which individual shape is most frequent in the Maura category without respect to color (an *independent-feature* strategy), then the likely inference is the new shape will be a diamond (most of the individual figures in the Maura category are diamonds). Murphy and Ross (2010a) obtained overwhelming evidence in their study in favor of the feature-conjunction strategy. Furthermore, this evidence was obtained even in experiments in which the researchers went at length to devise scenarios that would suggest to subjects that forms of feature independence should operate. Murphy and Ross (2010a) summarized the pattern of results by writing ". . . the conjunction strategy appears to be a heuristic that applies regardless of . . . statistical justification—and contrary to the widespread assumption that inductions are based on summary category-level information" (p. 8).

The exemplar-model prediction equations for the Figure 7 design (assuming that Maura is indeed the inferred target category) are as follows:
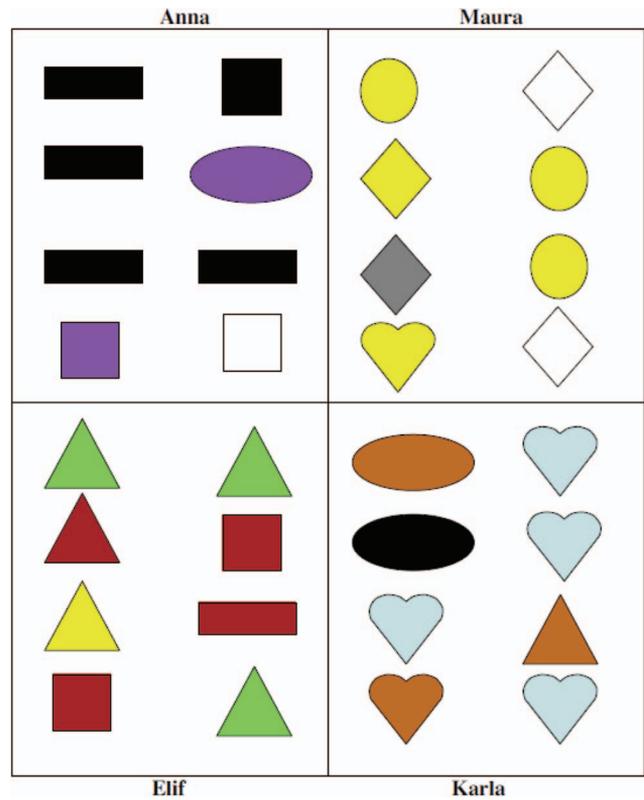

Figure 7. Experimental design from Murphy and Ross (2010a). From "Category versus object knowledge in category-based induction," by G. L. Murphy and B. H. Ross, 2010, *Journal of Memory and Language, 63,* pp. 1–17. See the online article for the color version of this figure. Copyright 2010 by Elsevier. Reprinted with permission.

Feature-conjunction

$$P(\text{circle}|\text{yellow}) = [3]/[(8CL) + (5 + 3C)$$

$$+ (L + 7CL) + (8CL)]. \quad (10a)$$

Independent feature

$$P(\text{diamond}|\text{yellow}) = [1 + 3C]/[(8CL) + (5 + 3C)$$

$$+ (L + 7CL) + (8CL)]. \quad (10b)$$

The denominators of Equations 10a and 10b are identical. And assuming typical parameter estimates (e.g., see Medin & Schaffer, 1978) for the similarity parameter C, the numerator of Equation 10a greatly exceeds the numerator of Equation 10b, so the prediction from the model is that P(circle | yellow) > P(diamond| yellow). Thus, the model correctly predicts Murphy and Ross's (2010a) robust finding in which subjects' inferences follow the predictions from the feature-conjunction strategy rather than the independent-feature strategy. Reversing this prediction would require large estimates of the similarity parameter C. These would arise only in situations in which subjects gave little attention-weight to the color dimension or in which the different colors were difficult to discriminate from one another.

Interpreting the Effects of Experimental Manipulations on Feature Inferences From Uncertain Categories

In each of the previous experiments reviewed in this article, a single robust result was reported that was predicted a priori by the proposed exemplar model across the vast range of its parameter space. By contrast, in the experiments that I consider in this section, there were sizable individual differences or the patterns of results changed with the associated experimental conditions. Therefore, I will not claim that the exemplar model predicts the results a priori. Instead, the issues are: (a) whether the observed results fall within the scope of the exemplar model, and (b) whether it can provide a natural interpretation of the results in terms of how parameter settings in the model may vary across individuals or changes in experimental conditions.

As noted at the outset of this article, the joint role of feature-based and category-based inference strategies has also been studied extensively by Hayes, Newell, and their colleagues. In one example, Griffiths et al. (2012) conducted experiments to test which conditions promote the use of feature-based versus category-based reasoning. Across two experiments they manipulated: (a) the coherence of the categories (i.e., whether the categories had high or low within-category similarity); (b) whether or not subjects were trained to classify category members prior to the feature-inference questions; and (c) whether or not subjects were provided with the specific exemplars on-screen during the feature-inference questions. In brief, Griffiths et al. found that subjects overwhelmingly used feature-conjunction strategies in conditions in which classification training did not take place (see also Griffiths, Hayes, Newell, & Papadopoulos, 2011), but there was a shift toward use of category-based independent-feature strategies when subjects did engage in classification training.

The essence of the feature-inference design used by Griffiths et al. (2012) is presented in tabular form in Table 2. (I do not report in the table values along a set of dimensions that were not relevant to the feature-induction task, which varied across the high-coherence and low-coherence conditions.) In the feature-induction task, the given feature was the Value 1 on Dimension 1, and the to-be-predicted feature was the value on Dimension 2. Given the Value 1 on Dimension 1, the inferred target category is the one denoted "Target"; a plausible alternative is denoted "Nontarget"; and the third is denoted "Irrelevant." Griffiths et al.'s design was used to contrast three major feature-inference strategies. According to a single-category independent-feature strategy, subjects would infer Value 1 on Dimension 2, because the majority of members from the target category have Value 1. According to a single-category feature-conjunction strategy, subjects would infer Value 4 on Dimension 2, because the majority of members from the target category that have Value 1 on Dimension 1 have Value 4 on Dimension 2. And according to a multiple-category feature-conjunction strategy, subjects would infer Value 5 on Dimension 2, because across *all* categories (i.e., not just the target category) most objects that have Value 1 on Dimension 1 have Value 5 on Dimension 2. As previewed above, under conditions in which subjects did not engage in classification training prior to inference, the evidence pointed toward single-category and multiple-category feature-conjunction strategies. But in conditions in which subjects did engage in classification training,

Table 2

Core Aspects of the Category Structure Used in Griffiths, Hayes, and Newell (2012)

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

Note. Dim. 1 = given feature; Dim. 2 = to-be-predicted feature.

there was a switch toward use of a single-category independent-feature strategy.

Letting S denote the similarity-mismatch parameter on Dimension 1, L the category-label mismatch parameter, and $D_i(j)$ the value j on Dimension i, the prediction equations from the exemplar model for this design are as follows:

Single-category independent-feature

$$P[D_2(1)|D_1(1)] = [3 + 6S]/[(10 + 6S) + (5L + 11SL) + (16SL)]. \quad (11a)$$

Single-category feature conjunction

$$P[D_2(4)|D_1(1)] = [5 + SL]/[(10 + 6S) + (5L + 11SL) + (16SL)].$$

(11b)

Multiple-category feature conjunction

$$P[D_2(5)|D_1(1)] = [2 + 5L + SL]/[(10 + 6S)$$

$$+ (5L + 11SL) + (16SL)].$$ (11c)

The denominators of Equations 11a–11c are identical, so the relative magnitudes of the feature-inference predictions depend only on the relative magnitudes of the numerators. These relative magnitudes are parameter dependent. In cases in which the given feature (on Dimension 1) is highly weighted (so that the magnitude of S is small) and the likely category is given little weight (so that the magnitude of L is big), the prediction is that the dominant inferences will be based on multiple-category feature conjunctions. If both the given feature and the inferred category are given high weight (so that the magnitudes of both S and L are small), then the dominant inference will be based on single-category feature conjunctions. And if the category-label feature is given high weight (small L), but the Dimension-1 feature is given low weight (big S), then the dominant inference will be in accord with the single-category independent-feature strategy.

Thus, one interpretation is that under the conditions in which subjects did not engage in prior classification training, they gave a great deal of attention to the given feature (and, therefore, to specific exemplars), yielding feature-conjunction behavior. By contrast, under conditions in which subjects engaged in extensive classification training prior to the feature-inference task, they gave less weight to the single given feature and relatively more weight to category-level information in making their inferences. I should note that, especially for the high-coherence category condition, many dimensions besides Dimension 1 were highly predictive of category membership, so subjects in Griffiths et al.'s (2012) experiment may indeed have learned to focus their attention on other dimensions besides Dimension 1 when given classification training. Thus, it is perhaps not too surprising that many of those subjects may have given little weight to that dimension during the feature-inference stage.

Another example of a stimulus design that yields parameter-dependent predictions from the exemplar model is one tested by Murphy and Ross (2010b) and Murphy, Chen, and Ross (2012). The core design is provided in Table 3 (although there were variants of this core design across different experiments). In this example, the given feature is square and at issue is the extent to which subjects will infer the most likely color to be vertical-stripe versus black. Given square, Federico is the most likely target category but Cyrus is a strong plausible alternative. If subjects focus only on the most likely target category, then the inferred color will be vertical-stripe; but if multiple categories are considered, then the inferred feature will likely be black. Under the usual testing conditions for this paradigm, although the most frequent inference tended to be the single-category one (vertical-stripe), a significant subset of subjects consistently gave the multiple-category inference (black). Furthermore, Murphy and Ross (2010b) discovered alternative testing conditions that had a dramatic effect on whether subjects

Table 3
Category Structure Used in Murphy, Chen, and Ross (2012)

Category	Exemplar no.	Shape	Color
Cyrus	1	Rectangle	Checkered
Cyrus	2	Rectangle	Black
Cyrus	3	Square	Black
Cyrus	4	Square	Black
Cyrus	5	Rectangle	Black
Cyrus	6	Square	Black
Cyrus	7	Square	Black
Cyrus	8	Rectangle	Checkered
Lindsey	1	Circle	White
Lindsey	2	Circle	White
Lindsey	3	Circle	Horizontal stripe
Lindsey	4	Heart	White
Lindsey	5	Heart	White
Lindsey	6	Circle	White
Lindsey	7	Circle	Horizontal stripe
Lindsey	8	Heart	Horizontal stripe
Lindsey	9	Heart	White
Lindsey	10	Circle	White
Federico	1	Square	Vertical stripe
Federico	2	Square	Black
Federico	3	Square	Black
Federico	4	Square	Vertical stripe
Federico	5	Square	Vertical stripe
Federico	6	Diamond	Wavy
Federico	7	Square	Vertical stripe
Federico	8	Square	Black
Monique	1	Heart	White
Monique	2	Triangle	White
Monique	3	Triangle	White
Monique	4	Heart	Dotted
Monique	5	Heart	White
Monique	6	Triangle	White
Monique	7	Triangle	White
Monique	8	Triangle	Dotted
Monique	9	Triangle	Dotted
Monique	10	Heart	White

tended to make the single-category or multiple-category choice. For example, whereas the usual procedure is to have subjects indicate only the target category, in a modified procedure subjects were asked to make probability judgments for all categories. Under these latter conditions, there was a dramatic increase in the number of inferences based on multiple categories (see Murphy et al., 2012 for an extensive investigation of factors that influence the strategy).

The exemplar-model predictions for the Table 3 design (assuming that Federico is the target category) are as follows:

Single category

$$P(\text{vertical} - \text{stripe}|\text{square}) = [4]/[(4L + 4SL) + (10SL)$$

$$+ (7 + S) + (10SL)].$$ (12a)

Multiple category

$$P(\text{black}|\text{square}) = [3 + 4L + 2SL]/[(4L + 4SL) + (10SL)$$

$$+ (7 + S) + (10SL)].$$ (12b)

The denominators of Equations 12a and 12b are identical. The relative magnitudes of the numerators are dependent on parameter

settings. Under conditions in which subjects give a great deal of weight to the category-label dimension (so that L is near zero), the model predicts that subjects will make the single-category inference. But if there is not a strong focus on the target-category label (so that the value of L is moderate to high in magnitude), then the model predicts that subjects will make the multiple-category inference. Thus, a reasonable interpretation is that under the conditions in which subjects are asked to make probability judgments for all four categories, information from those categories is not "zeroed out" and there is an increase in multiple-category inference.

Again, I emphasize that the example paradigms considered in this final section do not constitute a priori "tests" of the proposed exemplar model. Instead, the questions are whether any of the results falsify the model and whether the effects of the experimental manipulations can be understood in terms of changes in the model's parameter settings. It appears that the results can be well accommodated by the proposed model, but future research will be needed to judge whether the model provides a convincing and coherent story for these more complex designs.

The Rational Model

Although the present proposed model does a good job of handling the wide variety of results reported in these feature-inference paradigms, a natural question is whether alternative models may also capture the findings. As already noted, several important contenders fail to provide an adequate account of the results. Such models include: (a) multiple-category Bayesian models that integrate across independent-feature counts within the categories, (b) models that focus on *only* the single target category, regardless of whether the feature inferences are based on an independent count of features or on feature conjunctions, and (c) pure feature-conjunction models that assess feature conjunctions across the multiple categories without the target category having a special status. An important direction for future research will involve the development and testing of models from alternative frameworks that might be applied to the present tasks, including connectionist models of semantic knowledge (Rogers & McClelland, 2004), feature-based models of induction (Sloman, 1993), more general Bayesian models than the independent-feature version (e.g., Heit, 2000; Kemp & Tenebaum, 2009), and quantum-probability models (Busemeyer, Wang, & Lambert-Mogiliansky, 2009).

An alternative model that merits immediate consideration, however, is the *rational* model proposed by Anderson (1990, 1991), with important variants developed recently by Sanborn, Griffiths and Navarro (2010). The reason for particular interest in the rational model is that it is closely related to the independent-feature multiple-category Bayesian model that motivated many of Murphy and Ross's (1994) experiments. In brief, according to the rational model, observers are presumed to organize individual items into categories (or "clusters"). At any given stage of the learning process, the probability that an item joins a cluster is determined jointly by the prior probability of that cluster and by the degree of match between the individual features of the item and the individual features that compose the cluster. Importantly, the category label associated with an item acts as does any other feature and will also influence the clusters that are formed. When required to make a feature inference from a test probe, the observer is presumed to sum the probability of the feature given each cluster, weighted by the probability of the cluster given the test probe. In the special case in which the clusters that are formed correspond exactly to the experimenter's "intended" categories (as defined by the category labels), the rational model is essentially identical to the multiple-category independent-feature Bayesian model that was tested by Murphy and Ross. However, depending on parameter settings, the clusters that the rational model forms may not correspond to these intended categories. Thus, the predictions of the rational model with respect to Murphy and Ross's designs remain to be determined.

In a preliminary investigation of this question, I conducted simulations of the rational model for several of the designs considered in the present article. The details of these simulations are reported in the Appendix. For each design, I considered four variants of the model based on crossing two main factors. The first factor concerned parameter settings in the model. In his original applications, Anderson (1990) sometimes set the salience parameters for all features (including the category-label feature) at a default value equal to 1; whereas in other applications, he treated the feature saliences as free parameters. In this latter case, the estimates of the saliences of the perceptual features of the objects were roughly of magnitude 1, but the category-label parameter was near zero (.001), reflecting *high* salience—see Anderson (1990) for details. (Because the category-label feature is generally a special feature, it seems reasonable that it will often have a high-salience parameter estimate.) Thus, in the present investigations, I considered the qualitative predictions from the rational model for both types of parameter settings: In the first case, all salience parameters were set at 1; and in the second, the salience parameters of the perceptual features were set at 1, but the category-label salience was set at .001. The second factor concerned the detailed mechanism of cluster formation during the learning process. In Anderson's (1990) original scheme, at each stage of the cluster-formation process, a presented item joined the cluster that yielded the highest posterior probability (i.e., a max rule was used). Sanborn et al. (2010) considered some alternative algorithms for cluster formation. The one that they found to be the most successful was the same as Anderson's scheme, except items joined clusters in accord with a probability-matching rule rather than a max rule. I considered both the max and probability-matching versions here.

The results of these simulations for Experiments 1, 5, and 8 from Murphy and Ross (1994) are reported in Table 4. To review, in Experiment 1, Murphy and Ross compared feature inferences in an "increasing" condition to those in a "neutral" condition. They found no difference in feature inferences across these conditions. By contrast, as reported in Table 4, regardless of which of the four variants is applied, the rational model predicts that inferences for the queried feature should be greater in the increasing condition than in the neutral condition. In their Experiment 5, Murphy and Ross contrasted predictions from a "single-category" view with those of a "multiple-category independent-feature Bayesian" view: they found that observers gave greater probability judgments for the feature favored by the single-category view than for the feature favored by the multiple-category one. As reported in Table 4, the

Table 4

Rational Model Predictions of Feature-Inference Probabilities in Some Representative Experiments From Murphy and Ross (1994)

	Model version			
	Max rule		Matching rule	
Experiment number and feature inference type	$lab = .001$	$lab = 1$	$lab = .001$	$lab = 1$
1. P(black|triangle) [increasing]	.510	.489	.476	.411
1. P(white|square) [neutral]	.424	.435	.413	.382
5. P(black|triangle) [single-category]	.326	.405	.372	.374
5. P(white|triangle) [multiple-category]	.425	.458	.421	.412
8. P(vertical-stripe|circle) [correlated]	.511	.875	.603	.718
8. P(white|triangle) [uncorrelated]	.511	.483	.475	.428

Note. lab = category-label salience parameter value.

rational model again makes the wrong qualitative prediction, regardless of the model variant. And in their Experiment 8, Murphy and Ross found that the probability of inferences associated with correlated features (feature conjunctions) exceeded those associated with features that broke correlations. As reported in Table 4, for this experiment, the predictions from the rational model depend on the variant that is used. In particular, the variant that assumes high category-label salience and that uses a max rule for cluster formation predicts incorrectly that there should be no difference in inferences for the correlated versus uncorrelated features. (This variant of the model produces clusters that correspond with the "intended" categories.) But the other variants make the correct prediction, favoring the correlated features. The reason is that the other variants often form clusters that reflect the correlated perceptual features rather than clusters that reflect the category labels.

In sum, the predictions from important candidate versions of the rational model are often challenged by the results reported in Murphy and Ross's experiments. It is an open question whether reasonable modifications that maintain the spirit of the model would capture these challenging effects in parsimonious and convincing fashion.

Discussion

In summary, in this note I proposed and evaluated an explicit formal model to account for performance in the Murphy and Ross (1994) paradigm used for evaluating how people make feature inferences from specific exemplars and uncertain categorizations. The proposed model assumes that the inferred category label of a highly likely target category joins as an additional feature of an exemplar probe. Inferences to new features are then based on the machinery associated with standard exemplar models of categorization. I showed that the proposed model predicts a priori a wide variety of robust findings from this literature and can also be used to interpret the results from more complex designs in which patterns of behavior differ across individual subjects and associated experimental conditions. To my knowledge, this model is the first explicitly formalized one that accounts successfully for the diverse range of results that have been reported for the paradigm.

Various components of the model seem to be important in allowing it to capture the results. First, its exemplar-based memory representation and assumption of a nonlinear similarity rule give it sensitivity to specific exemplar information such as feature conjunctions and correlations. Second, the summed-similarity mechanism in its decision rule provides it with sensitivity to feature and exemplar base rates. Third, by allowing the inferred category label to join as a feature of an exemplar probe, the model's inferences can be strongly influenced by category-level information. Fourth, by allowing for attention-weighting of the different components of the exemplar probe, the model can describe how inferential decisions may vary across individual subjects and experimental conditions.

The version of the model that I evaluated here made the strong assumption that all subjects infer the category label of the most likely target category and that there is perfect memory for the specific exemplars that compose each category. One generalization would make allowance for multiple target-category inferences to be made, with predicted performance involving a probabilistic mixture across these multiple category inferences. A second generalization would make allowance for noisy memory representations of the stored exemplars. However, for the versions of the paradigm that were the focus of this note, these more complex generalizations were apparently not needed.

I should emphasize that the addition of the inferred category label to the probe is not a core assumption of exemplar models per se (although the idea has been advanced previously for related designs, e.g., Nosofsky et al., 2000). Instead, I propose it is a psychological process that may or may not accompany exemplar-based categorization and inference strategies. Under the conditions tested in the Murphy-Ross paradigm, the process may occur with high probability for several reasons. First, the experimenters often require the subjects to infer a category before making their other feature inferences. Second, the alternative categories are highlighted in separate spatial panels of the visual display, perhaps bringing category membership information into greater focus. Third, a single target category often clearly stands out as the one with the highest relative likelihood. However, the precise factors that govern whether the category label is added to the exemplar probe before other feature inferences take place remain to be determined in future research.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429. http://dx.doi.org/10.1037/0033-295X.98.3.409

Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Journal of Mathematical Psychology, 53,* 423–433. http://dx.doi.org/10.1016/j.jmp.2009.03.002

Griffiths, O., Hayes, B. K., & Newell, B. R. (2012). Feature-based versus category-based induction with uncertain categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 576–595. http://dx.doi.org/10.1037/a0026038

Griffiths, O., Hayes, B. K., Newell, B. R., & Papadopoulos, C. (2011). Where to look first for an explanation of induction with uncertain categories. *Psychonomic Bulletin & Review, 18,* 1212–1221. http://dx.doi.org/10.3758/s13423-011-0155-0

Hayes, B. K., & Newell, B. R. (2009). Induction with uncertain categories: When do people consider the category alternatives? *Memory & Cognition, 37,* 730–743. http://dx.doi.org/10.3758/MC.37.6.730

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology, 24,* 341–380. http://dx.doi.org/10.1016/0010-0285(92)90011-P

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7,* 569–592. http://dx.doi.org/10.3758/BF03212996

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116,* 20–58. http://dx.doi.org/10.1037/a0014282

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238. http://dx.doi.org/10.1037/0033-295X.85.3.207

Murphy, G. L., Chen, S. Y., & Ross, B. H. (2012). Reasoning with uncertain categories. *Thinking & Reasoning, 18,* 81–117. http://dx.doi.org/10.1080/13546783.2011.650506

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology, 27,* 148–193. http://dx.doi.org/10.1006/cogp.1994.1015

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). New York, NY: Cambridge University Press.

Murphy, G. L., & Ross, B. H. (2010a). Category vs. object knowledge in category-based induction. *Journal of Memory and Language, 63,* 1–17. http://dx.doi.org/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. http://dx.doi.org/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*(6), 869–874.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114. http://dx.doi.org/10.1037/0278-7393.10.1.104

Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. *Psychological Science, 2,* 416–421. http://dx.doi.org/10.1111/j.1467-9280.1991.tb00176.x

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7,* 375–402.

Papadopoulos, C., Hayes, B. K., & Newell, B. R. (2011). Noncategorical approaches to feature prediction with uncertain categories. *Memory & Cognition, 39,* 304–318. http://dx.doi.org/10.3758/s13421-010-0009-4

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

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review, 117,* 1144–1167. http://dx.doi.org/10.1037/a0020511

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280. http://dx.doi.org/10.1006/cogp.1993.1006

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133,* 166–188. http://dx.doi.org/10.1037/0096-3445.133.2.166

Appendix

Simulation Details for the Rational Model

The predictions from the rational model were averages computed across 1000 simulations of the model. Each simulation produced clusters by cycling through 12 learning blocks of the stimuli used in each experimental design. The order of the presented stimuli was randomized in each block. Following the cluster-formation process, the exemplar probes were presented to the model to compute its feature-inference probabilities for that simulation. Although the stimuli in Murphy and Ross's (1994) experiments were generally viewed in simultaneous visual displays, I assumed 12 blocks of sequentially presented stimuli in order to maintain comparability with the rational-model parameter values used in Anderson (1990). The rational model uses a "coupling" parameter that influences the computed prior probability of each cluster. Following Anderson (1990), I set the coupling parameter equal to .3 in all of the simulations. It should be noted that Nosofsky (1991) pointed out that the rational model is formally identical to the exemplar-based context model in cases in which the coupling parameter is set equal to zero during the cluster-formation process and then set at one when the observer makes classification decisions at time of transfer. If one made the additional assumption that the inferred category label is added to the exemplar probe when the observer makes feature inferences, then the rational model would reproduce the predictions from the exemplar model. However, these procedures would simply exchange the assumptions of one model for the other.