Title
Inferring Properties when Categorization is Uncertain: A Feature-Conjunction Account

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Hayes, Brett K.
Ruthven, Chris
Newell, Ben R.

Publication Date
2007

Peer reviewed

Inferring Properties when Categorization is Uncertain: A Feature- Conjunction Account

Brett K. Hayes (B.Hayes@unsw.edu.au)
Chris Ruthven (Cruthven@psy.unsw.edu.au)
Ben R. Newell (Ben.Newell@unsw.edu.au)
School of Psychology, University of New South Wales
Sydney, 2052, Australia

Abstract

Two studies examined how people make feature inferences about exemplars whose category membership is uncertain. Participants studied categorized exemplars, were given a feature of a novel item and asked to make predictions about other features of the item. Feature base rates were manipulated to test a series of possible inference strategies. Both studies found evidence for a feature conjunction strategy in which predictions are based on comparisons with exemplars similar to the test item. In both studies the majority of participants disregarded category membership when making feature predictions.

Keywords: Induction, Feature inference, Inductive reasoning

Introduction

Imagine that you were asked to predict the likelihood that the approach of the US government to the war in Iraq will change after the 2008 Presidential elections. It seems evident that to answer this question one needs to consider two possible outcomes (a Republican or a Democrat victory) and to think through the probabilities of policy change conditionalized on these outcomes. This sort of problem can be construed as one of making a property inference based on a category (the party affiliation of the next President) whose identity is uncertain. Most previous work on property inference has focused on predictions about the attributes of members of known categories (see Heit, 2000 for a review). Given the often unpredictable nature of the environment, however, it is inevitable that people will also make inferences about objects whose category membership has yet to be determined.

One approach to this problem, derived from Bayesian calculus and suggested in the earlier example, is to identify the possible categories to which an object might belong, derive the probabilities of the predicted property for each category, and then combine these conditional probabilities, weighting each according to the likelihood of the object being in that category. Bayesian models of human inference (most notably Anderson, 1991, but also see Tenenbaum, 1999) suggest that people do something like this when faced with property predictions under uncertainty. Arguing against this account, however, is a considerable body of evidence that people often ignore category uncertainty when making such predictions (see Murphy & Ross, 2007 for a review). Murphy and Ross (1994), for example, presented stimuli like those in Figure 1, said to have been drawn by different children. Participants were then shown novel instances with a given feature and asked to predict another feature, referred to as the "critical feature" (e.g., given that the object is a cross, how likely is it that it is brown?; given that the object is a diamond how likely is it that it is pink?). In each case one category is clearly the most likely or "target" category for the given feature (Liz drew the most crosses; Paula drew the most diamonds) but category membership is uncertain because instances containing the given feature are found in other categories (Paula and Jane also drew crosses). A Bayesian multiple-category approach leads to different estimates of the probability of the critical feature for these two examples (see Table 1 and Appendix for details). However, if people only consider the target category they should give similar estimates for each. Murphy and Ross (1994) found that most people followed the single-category strategy, even when they acknowledged that they were unsure about the category membership of test items. Such single-category reasoning has since been shown to be far more prevalent than the Bayesian alternative across a range of inference problems involving both artificial and natural categories (e.g., Malt, Ross, & Murphy, 1995; Murphy & Ross, 2005; Ross & Murphy, 1996).

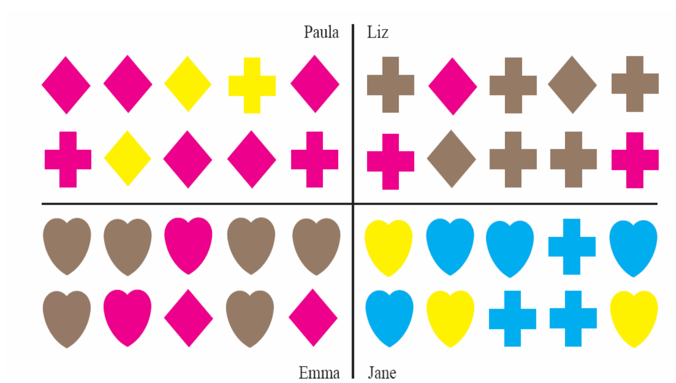

Figure 1: Example of a stimulus set used in Experiment 1.

These results are important because they show that people do not always follow Bayesian principles when making inferences about instances from uncertain categories. What is still not clear, however, is whether the single-category strategy that has been the focus of most of the previous work is the most prevalent non-Bayesian strategy that people employ when making uncertain inference. A serious problem with the psychological plausibility of both Bayesian and single-category approaches is that they assume that exemplar features are processed independently. In the single category strategy the given feature is used to identify the target category (e.g., in Figure 1, if a test item is cross-shaped Liz is the target). Inferences about the critical feature are then based on an assessment of relative feature frequencies within the target category *independent* of the frequency of the given feature. Seven out of Liz's ten drawings were brown so this should lead to a fairly confident prediction that it is the critical color. Note that the given feature of shape is only used to identify the target category, and thereafter plays no role in the feature prediction process.

A key assumption of many theories of object categorization (e.g., Love & Markman, 2003; Medin & Schaffer, 1978; Nosofsky, 1992), however, is that different feature dimensions are not treated as independent units. Instead people often encode the feature configurations in individual exemplars. Supporting this view are data showing that people are sensitive to feature correlations when making uncertain inferences (e.g., Murphy and Ross, 1994, Experiments 7-8).

This suggests a different way of dealing with the problem of making uncertain inferences - referred to as the "feature conjunction" approach. This involves focusing only on exemplars that possess a given feature and making property predictions based on the other features of these exemplars. If people follow this strategy they may ignore category-level information entirely. That is, they may base their feature predictions on the relative frequency with which different features are paired with the given feature across all available exemplars, ignoring category boundaries. According to this approach, after viewing Figure 1 someone who is told that an item is cross-shaped will examine all crosses in all available categories, and compare the relative frequency of different colors in these exemplars. The majority were brown so that would remain the critical feature. However, because more crosses are examined, and many of these are not brown, this would lead to a much lower probability estimate for this color than the single-category approach (see Table 1 and Appendix). In the same way, an experienced doctor who sees a patient with a livid rash on their chest and wants to predict other symptoms that the patient may have, might retrieve instances of previous patients with similar symptoms before assigning the patient to a diagnostic category (cf. Norman & Brooks, 1997).

The argument that induction when category membership is uncertain may be based on generalization from specific exemplars (rather than categories of exemplars) is consistent with exemplar models of categorization (e.g., Medin & Schaffer, 1978; Nosofsky, 1986). Such models propose that categories are represented by storing exemplars as individual memory traces, and classification occurs via comparison with these traces. In an uncertain induction task, the given feature would cue the retrieval of known exemplars that shared the feature. These could then be examined to discover which features co-occurred with the given.

The possibility that people use feature conjunction has been acknowledged in past work on uncertain induction (e.g., Murphy & Ross, 1994) but has not been examined in any detail. Murphy and Ross (1994, Experiments 5-6), for example, minimized the likelihood that this strategy would be employed by making sure that no feature was paired with the given feature more frequently than any other (i.e., feature conjunction did not yield a clear feature prediction). In other cases, the predictions of the feature conjunction approach have been confounded with those of other strategies. In Murphy and Ross (1994, Experiments 1-3; 2005), for example, a feature conjunction approach would have predicted the same findings as the single-category approach favored by the authors. Murphy and Ross' (2005) finding that probability estimates of a second feature given an observed feature were affected by manipulation of category validity but not cue validity is also consistent with feature conjunction.

The current studies therefore sought to advance understanding of how people make property inferences under category uncertainty by setting up situations in which the predictions of the feature conjunction approach were contrasted with those of category-based induction strategies. Experiment 1 compared feature conjunction with the single-category strategy suggested by Murphy and Ross (1994). Experiment 2 compared feature conjunction with a Bayesian multiple-category approach.

Experiment 1

This study compared the predictions of two strategies for making property predictions about instances whose category membership was uncertain. The first strategy was the single-category approach described by Murphy and Ross (1994) (which assumes that features are processed independently). The second was a feature conjunction approach in which predictions are based on examination of all exemplars that have the given feature.

Method

Participants

Twenty male and female employees of a market research firm (M_{age} = 25.1 years) were recruited by the second author.

Design and Materials

Two stimulus sets were constructed that followed the structure illustrated in Figure 1. In each set there were four categories containing ten exemplars that varied on two feature dimensions (shape and color). The cover story was that these were drawings done by different children. For each test item participants were given one feature of a novel exemplar whose category membership was unknown. They had to judge which category it was most likely to belong to and to predict what other feature it was most likely to have. For each stimulus set two types of inferences were generated. One of these was referred to as the neutral inference. For the stimulus set in Figure 1, an example of a neutral inference was predicting the color of an exemplar given that it was cross-shaped. The second type was referred to as the "increasing" inference (because some strategies predict a higher probability for the critical feature on these items than in the neutral case). An example from Figure 1 was predicting the color of an exemplar given that it was diamond-shaped. Stimuli were designed so that although different reasoning strategies always led to the same critical feature being predicted for any given item (in the above examples the expected or critical colors were "brown" and "pink" respectively), the estimated probability of this feature depended on the strategy being applied. Single-category reasoning led to very similar probability estimates for the critical features in neutral and increasing inferences. Feature conjunction led to higher probability estimates for the increasing inferences than for neutral items (see Table 1). The probability of identifying the target category was always the same for neutral and increasing inferences.

Table 1: Formal probability estimates for predictions of the critical feature based on different reasoning strategies[1]

	Strategy	Neutral items	Increasing items
Experiment 1	Single category	0.70	0.70
	Feature conjunction	0.38	0.67
Experiment 2	Single category	0.70	0.70
	Multiple category	0.40	0.40
	Feature conjunction	0.40	0.60

The frequency structure of these stimuli did not allow for a reversal of the direction of inferences (i.e., the probability estimates given in Table 1 did not hold for predictions about shape given color). Therefore all analyses were based on predictions of color given shape. However, filler items were developed with color given and a shape prediction required. Responses to these items were non-diagnostic with all strategies leading to similar probability estimates for critical feature predictions.

The feature frequency structure of the second stimulus set was identical to the first set but made use of different colors and shapes. The use of two sets meant that each participant responded to a total of two neutral and two increasing test items (and an equal number of filler items).

Procedure

The two stimulus sets were presented in random order. For each set participants answered questions about four test items, with items presented in random order. The first two questions for each item were about category membership (e.g., "*I have a drawing of a cross. Which child do you think is most likely to have drawn this? What is the probability that the child you just named drew it?"*). The purpose of these questions was to check that participants could identify the target category. These were followed by two property inference questions (e.g., *"What color do you think this cross is most likely to be? What is the probability that this cross is the color you just named?*"). All probability ratings were made on a 0 (Not at all likely) – 100 (Highly likely) scale. There was no time limit on task completion.

Results and Discussion

Preliminary analyses established that feature predictions did not vary across stimulus sets. All subsequent analyses were collapsed across this factor. Data from one participant were dropped from the analyses because the partcipant gave ceiling-level probability estimates for all category judgments and inferences.

Strategy predictions were based on the assumptions that people could readily identify the target category given the test feature, and that they recognized that category membership was uncertain. To check these assumptions we calculated the proportion of items where the target category was correctly identified and the probability estimates attached to these judgments. The target category was always identified correctly for neutral and increasing items. The mean probability estimate for these judgments was 0.64 across item types. People had no difficulty in identifying the category to which a test exemplar was most likely to belong, but recognized the uncertain nature of this judgment.

When people were asked to predict an additional exemplar feature after being given either its shape or color they always selected the critical feature that was predicted by the respective reasoning strategy for both neutral and increasing items. Our main interest therefore was in the probability estimates attached to feature predictions. Estimates given for increasing items (M = 0.693, SD = 0.212) were significantly above those given for neutral items (M = 0.612, SD = 0.168), $t(18) = 2.215$, $p < .05$. This suggests that many people were using a feature conjunction strategy to make feature predictions. Looking at individual response patterns, ten participants gave higher estimates (i.e., greater than a 3% difference) for increasing than for neutral items, suggesting that they used feature conjunction. Six gave similar estimates for neutral and increasing items, and three gave higher estimates for the neutral items.

This is the first positive evidence that some people employ a feature conjunction strategy rather than a categorical strategy when making feature predictions under conditions of uncertain category membership. Contrary to Anderson's (1991) rational model and Murphy and Ross' (1994) single-category account, these findings suggest that when people make predictions about exemplars with uncertain category membership many disregard category bounds. Instead predictions may be based on an

[1] Although formal estimates are derived we did not assume that intuitive probability ratings would approximate these values. Like Murphy and Ross (1994) our interest was only in whether people gave different estimates for neutral and increasing items.

examination of individual exemplars from multiple categories that have a given feature. The results also show that a majority of participants do not treat feature dimensions as independent when making inductive inferences.

Before we can be certain that we have found firm evidence of the use of feature conjunction, however, we need to consider at least one other possible explanation. When constructing the stimuli it was difficult to produce divergent predictions for the single-category and feature prediction accounts while holding constant the predictions of all other strategies. In particular, the Bayesian multiple-category approach also predicted a higher probability of the critical feature for increasing items than for neutral items (see Appendix). Although there is little previous evidence that people use this strategy (cf. Malt, et al., 1995; Murphy & Ross, 1994, 2005), to be confident in our conclusions about feature conjunction it was important to remove this confound. This was the main aim of Experiment 2.

Experiment 2

Experiment 2, therefore, set up a direct comparison of feature conjunction and Bayesian multiple-category reasoning. The experimental design was similar to Experiment 1 except that stimuli were constructed to contrast predictions based on these two strategies, while controlling for other possible strategies (e.g., single-category reasoning). The formal probability predictions derived from these strategies are summarized in Table 1. The feature conjunction approach predicted different intuitive probability estimates for neutral and increasing items, whereas the multiple-category approach predicted no difference between these estimates.

A subsidiary aim of this study was to extend our findings to a wider range of stimulus formats. First, we constructed parallel sets of shape/color figures so that participants were given test items that required predicting exemplar color given shape *and* shape given color. Second, we developed alternative stimulus sets using non-geometric stimuli. Love and Markman (2003) have suggested that people do not typically regard object shape and color as independent attributes, but instead treat color as a predicate for shape. Such stimuli would seem particularly amenable to a strategy like feature conjunction strategy that focuses on the configurations of individual exemplars. As a more stringent test of the feature conjunction strategy therefore we included exemplars that were composed of feature dimensions that we believed would be less likely to be treated as integral configurations (i.e., number-letter combinations).

Method

Participants

Forty two male and female employees of a market research firm (M_{age} = 26 years) were recruited by the second author. None of these took part in the previous study.

Design and Procedure

The design and procedure followed Experiment 1 with three important exceptions. First, neutral and increasing test items were designed to compare the predictions of feature conjunction with Bayesian multiple-category reasoning. An example is given in Figure 2. A neutral inference for this set was predicting an exemplar's color given that it was a square (critical feature = purple). An example of an increasing inference was predicting exemplar color given that it was a circle (critical feature = red). The multiple category approach always predicted similar probability estimates for these two types of items, whereas feature conjunction predicted a higher estimate for increasing items (see Table 1).

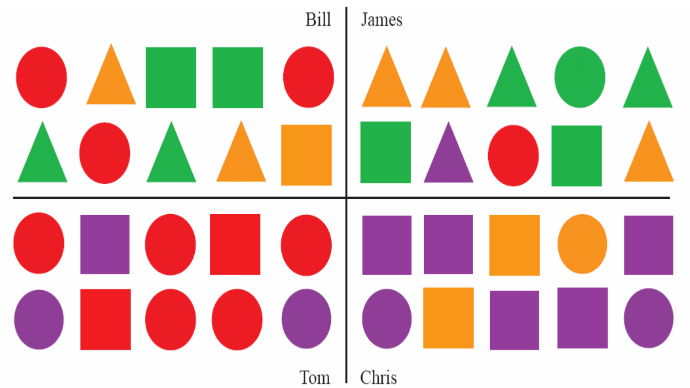

Figure 2: Example of a geometric shape/color set used in Experiment 2

The second innovation was the construction of parallel stimulus sets that had the same feature frequency structure as the shape/color stimuli illustrated above but which called for predictions of object shape given object color. Finally we created alternative stimulus sets made up of number-letter combinations (see Figure 3 for an example). The statistical structure of these sets was identical to the geometric items. Like the geometric items sets were constructed that called for the prediction of a number given a letter or vice versa. The order of appearance of letters/numbers in category exemplars always matched the direction of prediction to be made (e.g., when numbers appeared first, followed by letters, participants were given a number at test and asked to predict the corresponding letter).

Participants were randomly allocated to geometric or number-letter conditions. For the number-letter format the cover story was that the combinations were codes used by four different spies. After being given one component of a novel code participants were asked to judge which spy produced it and to predict the second code component. In each condition participants completed four test items (two neutral, two increasing) and four nondiagnostic fillers (which were not analyzed). As in Experiment 1, participants were asked to first decide on the category membership of the test item given a feature, then to predict the feature most likely to be found in the item, and estimate its probability. The direction of prediction (color given shape vs. shape given color or letter given number vs. number given letter) was counterbalanced across items.

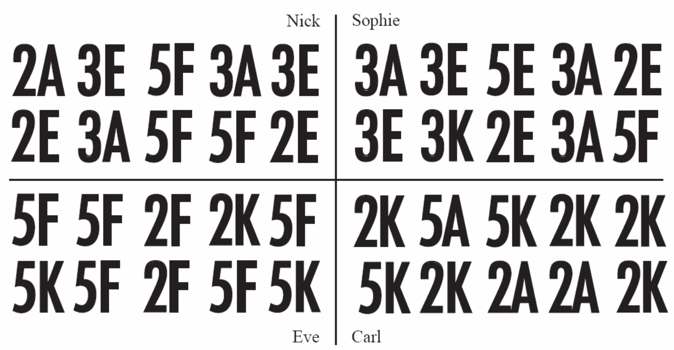

Figure 3: Example of a number-letter set (NB: Feature frequency structure is identical to Figure 2).

Results and Discussion

Preliminary analyses established that feature predictions in the two task formats were unaffected by the direction of the prediction so all subsequent analyses were collapsed across this factor.

The target category was always identified correctly for both neutral and increasing items. The mean probability estimate for these judgments was 0.62 across stimulus formats and item types, again indicating that participants recognized the uncertainty regarding category membership.

The critical feature was predicted for 98% of items. Probability estimates were only analyzed for predictions of this feature. Estimates were entered into a 2 (inference type: neutral vs. increasing) x 2 (stimulus format) ANOVA with repeated measures on the first factor. People generally gave higher probability estimates for predictions about increasing items (M = 0.641, SD = 0.130) than neutral items (M = 0.590, SD = 0.161), $F(1, 40) = 4.526$, $p < .05$. Format had no overall affect on estimates and did not interact with inference type. At the individual level seventeen participants gave higher estimates for increasing items, consistent with feature conjunction. Sixteen gave similar estimates for neutral and increasing items, and nine gave higher estimates for neutral items.

These data show that across both stimulus formats a narrow majority of participants used a feature conjunction approach when making property predictions under uncertainty. This also suggests that the evidence favoring the feature conjunction account in Experiment 1 was not due to the use of a multiple-category strategy.

General Discussion

These studies were concerned with how people make inferences about exemplar features when an exemplar's category membership is uncertain. Previous work on this issue has suggested two possible reasoning strategies, a Bayesian multiple-category strategy (Anderson, 1991) and a single-category strategy (Murphy & Ross, 1994), with the balance of evidence favoring the latter approach. Our studies suggest that the answer to the question of how people make inferences under category uncertainty is more complex. In both of our studies a substantial proportion of people did not use category-level information as a basis for making inferential predictions. Instead many employed a feature-conjunction strategy, basing predictions on an examination of instances that contained the feature given during the inference test. This approach is non-categorical in that all exemplars with the given feature are considered when making predictions, regardless of their category membership. Experiment 2 showed that this strategy is applied across a variety of stimulus formats. These results are particularly impressive given that inferences were always made after participants had identified the category that was most likely to contain the test instance. So even though participants were aware of the categorical structure of the stimuli, and could readily identify the category that was most likely to be associated with a test item, they used exemplars from all available categories when making feature predictions.

Our findings support the view that people do not necessarily treat feature dimensions as independent when doing categorization or induction (cf. Medin & Schaffer, 1978). Instead, we found that, after observing a feature of a novel exemplar, many people made inferences by focusing only on exemplars that had this feature.

Our argument that feature inference may sometimes operate by direct comparison with known exemplars, regardless of their category membership, is novel. In many respects, however, this approach is consistent with the general features of exemplar models of categorization. Such models do not store summary information about feature relations within or between categories. Instead, judgments about these relations, such as whether features are correlated, are based on the retrieval and examination of relevant exemplars. In the same way, when the category membership of a stimulus is uncertain, inferences about its other features may be based on direct comparison with similar instances retrieved from memory, without a preliminary decision about category membership.

These data suggest that single-category and Bayesian multiple-category models underestimate the variety and complexity of the strategies that people use when asked to make inferences without certain category membership. We have shown that, in addition to categorical strategies, people may employ exemplar-based or associative approaches. An important challenge for future work in this field is to examine the factors that lead people to adopt different reasoning strategies. One factor that is likely to be important is the perceived relevance of the category to the property that is being predicted (cf. Ross & Murphy, 1996). Categories are more likely to influence predictions when they are seen as particularly informative or relevant (e.g., knowing which party is governing when predicting foreign policy). In contexts where category structures are more arbitrary, however, people may eschew their use in favor of direct comparison with individual exemplars.

Acknowledgments

This work was supported by Australian Research Council Discovery Grant DP0770292 to the first and third authors. We thank Brian Ross for his comments on this work.

References

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

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

Love, B. C. & Markman, A. B. (2003). The non-independence of stimulus properties in human category learning. *Memory and Cognition, 31*, 790-799.

Malt, B. C., Ross, B. H., & Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 646-661.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-308.

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

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

Murphy, G. L., & Ross, B. H. (2007). Use of single or multiple categories in category-based induction. In A. Feeney & E. Heit (Eds.) *Inductive Reasoning: Cognitive, Mathematical, and Neuroscientific Approaches*. Cambridge University Press.

Norman, G. R., & Brooks, L. (1997). The non-analytical basis of clinical reasoning. *Advances in Health Sciences Education, 2*, 173-184.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Nosofsky, R. M. (1992). Exemplars, prototypes and similarity rules. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.) *From Learning Theory to Connectionist Theory: Essays in Honor of W. K. Estes* (Vol. 1, pp. 149-168). Hillsdale, NJ: Erlbaum.

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

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. *Advances in Neural Information Processing Systems, 11*, 59-68.

Appendix

Experiment 1: Example probability calculations (for Figure 1).

For the Neutral item the given feature was *cross* and the critical feature was *brown*. Liz was the target category. For the Increasing item the given feature was *diamond* and critical feature was *pink*. Paula was the target.

Feature Conjunction. Across all categories there were 13 crosses. Five were brown, so $p(Brown| Cross) = 0.38$.
There were 12 diamonds. Eight were pink so
$p(Pink| Diamond) = 0.67$

Single-Category. Seven of Liz's ten of drawings were brown so $p(Brown| Cross) = 0.7$
Seven of Paula's ten drawings were pink so $p(Pink| Diamond) = 0.7$

Multiple-Category. Liz was the target, but Paula and Jane also drew crosses (total = 13). Applying Bayes' formula for predicting feature *j* given observed feature *F* across *k* categories:

$$p(j|F) = \sum_k p(k|F)\, p(j|k)$$

$p(Brown| Cross) = ((7/13)\cdot(7/10)) + ((3/13)\cdot(0/10)) + ((0/13)\cdot(6/10)) + ((3/13)\cdot(0/10)) = 0.37$

Although Paula was the target, Liz and Emma also drew diamonds (total = 12). Applying the formula:
$p(Pink| Diamond) = ((7/12)\cdot(7/10)) + ((3/12)\cdot(3/10)) + ((2/12)\cdot(4/10)) + ((0/12)\cdot(0/10)) = 0.55$

Experiment 2: Example probability calculations (for Figure 2).

For the Neutral item the given feature was *square* and the critical feature was *purple*. Chris was the target. For the Increasing item the given feature was *circle* and critical feature was *red*. Tom was the target.

Feature Conjunction. There were 15 squares. Six were purple so $p(Purple| Square) = 0.40$.
There were 15 circles. Nine were red so
$p(Red| Circle) = 0.6$

Single-Category. Seven of Chris' ten drawings were purple so $p(Purple| Square) = 0.7$
Seven of Tom's ten drawings were red so $p(Red| Circle) = 0.7$

Multiple-Category. Chris was the target, but Bill, James and Tom also drew squares (total = 15). Applying the formula:
$p(Purple| Square) = ((7/15)\cdot(7/10))+((3/15)\cdot(3/10)) + ((2/15)\cdot(1/10)) + ((3/15)\cdot(0/10)) = 0.40$

Tom was the target but Bill, James, and Chris also drew circles (total = 15). Applying the formula:
$p(Red| Circle) = ((7/15)\cdot(7/10))+((3/15)\cdot(0/10)) + ((2/15)\cdot(1/10)) + ((3/15)\cdot(3/10)) = 0.40$