A Comparison of Three Measures of the Association Between a Feature and a Concept

Matthew D. Zeigenfuse (mzeigenf@msu.edu)
Department of Psychology, Michigan State University
East Lansing, MI 48823 USA

Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, University of California, Irvine
Irvine, CA 92697 USA

Abstract

In this paper, we develop three measures of association between concepts and features from three measures of category structure preference. These measures are total cue validity, feature possession score, and category utility. We compare these measures experimentally using stimuli from the Leuven Natural Concept Database (de Deyne et al., 2008). We find the measure developed from feature possession score, collocation, best captures human associations between concepts and features, followed by the measure developed from category utility, and finally the measure developed from total cue validity. We discuss how these results can be applied to open questions in categorization and similarity judgment such as domain differences in representation and modeling the effects of context.

Keywords: Mental representation; categorization; similarity judgment

Suppose someone asks you which of the features *has wings* and *sings* you associate more with the concept *birds*. What about the relationship between the two features and *birds* determines your answer? One possibility is that it is determined by how many birds have each feature. Almost all birds have wings but only a few can sing, so you would associate *has wings* with *birds* more than *sings*. Another possibility is that it is determined by how specific the feature is to birds. A number of animals besides birds have wings, for instance bats and wasps, but no other animals sing, so you would associate *sings* with *birds* more than *has wings*.

Associations between concepts and features form the basis of many theories of mental representation. For example, feature-based theories of mental representation hold that a concept is represented in terms of some number of (normally discrete) features. The associations people form between concepts and features determine which features will be used to represent a particular concept. Alternatively, dimensional theories of representation hold that a concept is represented by its values on some number of continuous dimensions. People's associations determine which dimensions will be used to represent a concept and what the values on each dimension will be.

The purpose of this paper is to compare empirically three different measures of feature association using their predictions of which feature you would choose. We procedes as follows. In the first section, we develop the three measures of feature association from measures predicting people's preferences for particular category structures. In the second, we describe an experiment comparing the three measures. In the third, we present the results of this experiment. Finally, we discuss what these results mean for other areas of cognitive psychology such as categorization and similarity judgment.

Association Measures

Each of the category structure preference measures used here were originally developed to explain basic level preference. Within a category hierarchy, the basic level is the preferred level of abstraction for describing objects. Key findings are objects are categorized into basic-level categories more quickly than sub- or super-ordinate categories, basic level objects are named faster, objects are described preferentially with basic level names, more features are listed at the basic level than at the superordinate level, basic level names are learned before names at other levels, and basic level names tend to be shorter (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

How basic people consider a category to be is a measure of their preference for its category structure. Since people prefer basic level categories to non-basic level categories for a large number of mental operations related to categorization, it seems reasonable to think of these categories as the standard for what people prefer categories to look like. Consequently, as how basic a category is considered to be decreases, so does people's preference for its category structure.

We derive measures of association between a feature and a concept for each of three measures of category structure preference: total cue validity (Rosch et al., 1976), feature possession score (Jones, 1983), and category utility (Corter & Gluck, 1992). To do this, we apply a principle we term the *feature impact hypothesis*. It states that the features people associate with a concept should be the features that have the largest impact on category structure preference for the category consisting of the concept's instances. Categories are abstracted from objects at least in part to help infer whether a novel object has one or more unobserved features given one or more features that have already been observed. Thus, it stands to reason that those features that caused a person to prefer a particular category structure in the first place should be the same features that the person now associates with the category.

The remainder of this section develops three measures of the association between a feature and a category. For each of these measures – cue validity, collocation, and change in category utility – we start by defining the measure of category structure on which the measure is based. Respectively, these are total cue validity, feature possession score, and category utility. We then apply the principle that people associate high impact features with categories in order to derive the three measures of association.

Cue Validity

Cue validity is the conditional probability that a stimulus belongs to category c given that it possesses feature f_k. Intuitively, it can be thought of as the degree to which having feature f_k implies being a member of category c. Cue validity is high when most of the stimuli that possess feature f_k are members of c and is low when most of the stimuli are not members of c.

Rosch et al. (1976) suggest using cue validity to determine which categories belong to the basic level. They argue that some of the features used to represent a stimulus will be linked with category c and propose using the sum of the cue validities of the linked features to determine which category is the basic level category. Unfortunately, Rosch and colleagues do not offer a method for determining which features are linked with which categories.

Applying the feature impact hypothesis to total cue validity leads to each feature's cue validity being its degree of association with the category. We would like to find those features that offer the largest individual contributions to the total cue validity. The contribution of an individual feature that is linked with the category simply its cue validity. Though the original theory developed by Rosch et al. (1976) does not specify criteria for selecting which features to link with a category, it seems reasonable to assume that a feature is more likely to be linked with a category when its cue validity with the category is high. Then, features with high cue validity have the most impact on category structure preference, since these features are most likely to be linked with a category and have the most impact when linked.

Collocation

Jones (1983) defines feature possession score in terms of another quantity called *collocation*. The collocation of feature f_k is

$$p(c|f_k) \times p(f_k|c), \tag{1}$$

where $p(c|f_k)$ is the product of the cue validity of f_k, and $p(f_k|c)$ is the conditional probability of observing f_k in a stimulus given that the stimulus is a member of category c. Intuitively, collocation can be thought of as the degree of bi-implication between f_k and c. Collocation is high for features that are possessed by most members of c but few non-members. From these individual collocations, feature possession score is computed by associating each feature to the category it has the largest collocation with and summing the collocations of all features associated with c.

Applying the feature impact hypothesis to feature possession score leads to each feature's collocation being its degree of association with the category. As with total cue validity, we would like to find those features that offer the largest individual contributions to feature possession score. Since collocation is only large when most exemplars of a category have a feature but most exemplars of different categories do not, collocation can only be large for a single category. Then, when a feature's collocation is large for a particular category, the feature will have a large impact on preference for that category's structure.

Category Utility

Category utility (Corter & Gluck, 1992) derives from the assumption that the purpose of a category is to convey information about the features of its exemplars. It is defined as the overall increase in the probability of correctly guessing, using probability matching, whether an object has each member of a set independent features upon learning that the object belongs to category c, normalized by the probability of observing c. Probability matching means that people guess whether an object has a feature with probability equal to the probability that the object has the feature. For example, if $p(f_k)$ is the probability an object has feature f_k, then $p(f_k)$ is the probability a person will guess the object has feature f_k and $[p(f_k)]^2$ is the probability they will correctly guess the object has feature f_k. Then, for a set of M independent features, the category utility is

$$p(c) \sum_{k=1}^{M} \{[p(f_k|c)]^2 - [p(f_k)]^2\}. \tag{2}$$

Applying the feature impact hypothesis to category utility leads to each feature's increase in probability of having its presence correctly guessed,

$$[p(f_k|c)]^2 - [p(f_k)]^2, \tag{3}$$

being its degree of association with the category. As with the previous two structure preference measures, we would like to find those features that offer the largest individual contributions to category utility. Equation (2) shows that those features that offer the greatest increase in being correctly induced will have the most impact on category utility.

Methods

We describe in this section an experiment to determine which of the three association measures best captures human associations between concepts and features. Overall, the experiment consists of presenting a number of triads containing a concept and two features and asking participants to select which of the two features they more strongly associate with the concept.

Each triad was constructed to differentiate two measures of association. We selected feature pairs such that one measure, measure A, finds the first feature strongly associated with the concept but not the second and the other measure, measure B, finds the second feature strongly associated with the concept but not the first. Then, a participant selecting the first feature supports measure A being the better model of human associations, and a participant selecting the second feature supports measure B the better model.

Recall our introductory example with the concept *birds* and the features *has wings* and *sings*. The probability that an animal has wings greatly increases upon learning that it is a bird, but the probability that an animal sings only modestly increases. Because of this, category utility judges the association between *birds* and *has wings* to be high, and the association between *birds* and *sings* to be low. Alternatively, bats and a number of insects have wings, but no animals other than birds sing. Cue validity then judges the association between *birds* and *has wings* to be low, and the association between *birds* and *sings* to be high. Thus, selecting *has wings* as being more associated with *birds* supports category utility being a better model of human associations than cue validity, and selecting *sings* supports cue validity being a better model than cue validity.

Our description of the experiment consists of three parts. First, we describe the process used for creating our stimuli. We then describe the participants in our experiment. Finally, we describe the experimental procedure.

Stimuli

The stimuli used in our experiment consist of triads consisting of a concept and two features. To choose the triads we start with the feature representations of 129 animals in terms of 765 features and 166 artifacts in terms of 1295 features contained in the Leuven Natural Concept Database (de Deyne et al., 2008), a database of normative data for a large number of animal and artifact concepts. The 129 animals are distributed among 5 animals categories, *mammals*, *birds*, *fish*, *insects* and *reptiles*, and the 166 artifacts are distributed among 6 artifact categories, *clothing*, *kitchen utensils*, *musical instruments*, *tools*, *vehicles* and *weapons*. These 11 categories are the concepts in our experiment. Since we know which category each stimulus belongs to, we can use these stimuli to compute the association between each feature and category using each of the three measures.

To illustrate this process, we will compute the association between the concept *birds* and the feature *has wings*. Recall that cue validity feature f_k is the probability a stimulus belongs to category c given the stimulus has f_k. In our case, this means that the cue validity of the feature *has wings* for the concept *birds* is

$$\frac{\#\text{ birds with wings}}{\#\text{ animals with wings}}.$$

The Leuven data set contains 43 stimuli that have wings, of which 30 are birds, so this cue validity is $30/43 = 0.69$.

Now recall that Equation (1) gives the collocation of feature f_k as $p(c|f_k) \times p(f_k|c)$ for category c. In our case, this is the cue validity of *has wings* times the proportion of *birds* exemplars with *has wings*,

$$\frac{\#\text{ birds with wings}}{\#\text{ animals with wings}} \times \frac{\#\text{ birds with wings}}{\#\text{ birds}}.$$

The Leuven data set contain 30 birds exemplars in total, so the collocation of *has wings* is $30/43 \times 30/30 = 0.69$ for *birds*.

Finally, Equation (3) gives a feature f_k's contribution to category utility as $[p(f_k|c)]^2 - [p(f_k)]^2$. In our case, $p(f_k)$ is the proportion of animals which have the feature *has wings*, feature f_k's contribution to category utility is

$$\left(\frac{\#\text{ birds with wings}}{\#\text{ birds}}\right)^2 - \left(\frac{\#\text{ animals with wings}}{\#\text{ animals}}\right)^2.$$

The Leuven data set contains 129 animals exemplars, 43 of which have the feature *has wings*. Then, the contribution of *has wings* to category utility is $[30/30]^2 - [43/129]^2 = 0.88$.

Once we had computed the association between each five animals concepts and 765 animals features and each of the six artifacts concepts and 1295 artifacts features, we needed to select triads of one concept and two features to present to participants. We had two primary goals in mind when doing this. First, we wanted to select triads whose features have the largest possible difference in predicted association between the two measures being compared. Second, we wanted to represent as many different concepts and features as possible. Whenever possible we used each concept in at least one comparison for each set of measures. This resulted in 61 triads of one concept and two features.

Participants

Our participants consisted of thirteen students and faculty from the University of California, Irvine. Of these thirteen, one was faculty, two were undergraduates, and ten were graduate students. Their ages ranged from 20 to 42. No compensation of any kind was offered.

		% First Measure Chosen		
Measure 1	Measure 2	Animals	Artifacts	All
Collocation	Cue Validity	94	90	92
Collocation	Category Utility	52	72	62
Category Utility	Cue Validity	83	70	77

Table 1: Percentages of comparisons in which participants chose the feature whose highest association was from by the first model. The first two columns give the measures being compared. The last three columns give the percentage of comparisons in which Measure A was chosen for animals, artifacts, and all comparisons, respectively.

Procedure

Stimuli for this experiment were presented to participants in MATLAB. The format for each was identical to the following.

```
Which of these features best describes
BIRDS:
(1) HAS WINGS
(2) SINGS

Response (Enter 1 or 2):
```

In these stimuli, the concept always comes in the same position as *birds* in the example. The features always following on separate lines with the number 1 preceding the first feature and the number 2 preceding the second. Participants were always reminded to enter 1 if they preferred the first feature and two if they preferred the second.

In performing the experiment, each participant first viewed a set of instructions explaining how to complete this experiment. Participants indicated they had read the instructions, at which point, the experiment started. The 61 triads were shown in two segments. In the first segment, all of 32 animals triads were shown. In the second, all the 29 artifacts triads were shown. The presentation order of the triads was randomized across participants, as was the which of the two features was presented first. Finally, the display was cleared after each response, so that only one stimulus was ever displayed at a time.

Results

In this section, we present the results of our experiment. This is done in two parts. In the first part, we look at how often the high association feature for each measure was preferred in comparisons with the two other measures across all participants. In the second, we look at these results on an individual basis.

Overall

Table 1 shows how often participants preferred features associated with each measure in comparisons. The first two columns of the table show the measures being compared. The last three columns show the percentage of comparisons in which participants chose the feature whose highest association was assigned by the measure in the 'Measure 1' column. The 'Animals', 'Artifacts', and 'All' columns show these percentages for animals, artifacts, and all (animals and artifacts together) comparisons. A percentage above 50 indicates that Measure 1 accounted for people's associations better than Measure 2. A percentage below 50 means the reverse. For example, the '94' in the 'Animals' column of the first row means that the feature with high collocation and low cue validity was chosen over the feature with high collocation and low cue validity in 94% of comparisons. This indicates that collocation accounted for people's choices better than cue validity.

The percentages in Table 1 give an overall ordering on how comparatively well the three measures fit human judgments of association. It shows that collocation best describes human judgments, followed by category utility, and finally cue validity. This ordering is consistent across domains.

Individual

Figure 1 shows how often each of the three association measures was chosen by each of our participants in comparisons. In the plots, solid lines correspond to choices in comparisons between features of animals concepts and dashed lines correspond to preferences in comparisons between features of artifacts concepts. As in Table 1, the dependent measure is the percentage of comparisons in which the first of the two measure in each title was chosen. For example, Participant 4 always chose the feature with high collocation in comparisons between cue validity and collocation . Both the solid and dashed lines are zero for Participant 4 in the first plot, indicating that in comparisons between cue validity and collocation, cue validity was chosen in 0% of comparisons.

The plots in Figure 1 allow us to draw a number of conclusions about the presence of differences across individuals and domains. The first plot shows each participant's choices in comparisons involving cue validity and collocation. Here we find that participants consis-

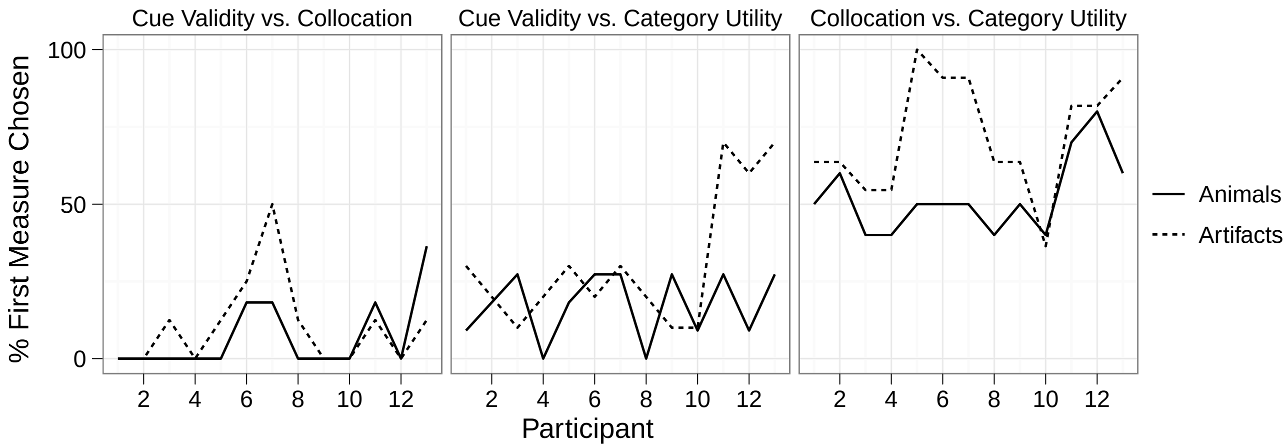

Figure 1: Percentages of comparisons in which each participant chose the feature whose highest association was from the first model. Individual choices for each pair of compared measures are shown in a different plot. In each of these plots, the independent axis is participant number and the dependent axis is the percentage of times the feature given high association by the first model was chosen. Values for animals comparisons are indicated by the solid line. Values for artifacts comparisons are indicated by the dashed line.

tently chose the feature with high collocation regardless of whether the concept was an animal or artifact.

The second plot shows participant's choices in comparisons involving cue validity and category utility. In this plot, Participants 1 − 10 consistently chose the feature with high category utility regardless of whether the concept was an animal or artifact. Participants 11 − 13, however, chose the feature with high category utility for most animal comparisons, but the feature with high cue validity for most artifacts comparisons. This suggests that there may be individual or domain differences in which of these measures best captures people's choices.

Finally, the third plot shows participant's choices between features favored by collocation and category utility. Here, Participants 2, 11, 12 and 13 consistently chose the feature with high collocation regardless of whether the concept was an animal or artifact, and Participant 10 consistently chose the feature with high category utility. Participants 1, 5, 6, 7 and 9 were chose the features with high collocation and category utility in equal proportions for animals comparisons, but most often chose features with high collocation in artifacts comparisons. Participants 3, 4, and 8 chose features with high category utility most often in animals comparisons, but features with high collocation most often in artifacts comparisons. The large variety in participant's choices strongly suggest individual or domain differences in which of collocation and category utility best captures people's choices.

Discussion

Our results have implications for many related fields of cognition. In this section, we discuss three such implications. First, we relate our finding of potential domain differences in which association measure best captures human associations between concepts and features to the previous findings of domain differences in the way concepts are represented (e.g. Keil, 1989; Zeigenfuse & Lee, 2010). Second, we suggest how our results could be used to understand how context affects categorization and similarity judgments. Finally, we relate our work to work by Murphy (1982) arguing against cue validity as a measure of category basicness.

Domain Differences

Numerous authors have found that the types of features used to represent concepts depends upon the domain to which those concepts belong. Zeigenfuse and Lee (2010), for example, looked at the ability of sets of features selected according to collocation to fit human similarity judgments. They found that features selected in this way fit similarity judgments between animals considerable better than between artifacts. This suggests that people may use different criteria determine a feature's importance depending on the domain of the stimulus.

We also find that different domains of exemplars are may be represented in fundamental different ways. The marked difference between how often people chose features with high collocation over features with high category utility in comparions among features of animals concepts versus artifacts concepts suggests that they rate importance more like collocation for artifacts than animals. Since the Zeigenfuse and Lee (2010) heuristics select feature sets according to their each feature's importance, this suggests that they may be selecting features in a manner more similar to collocation for artifacts than animals.

Context

Context has been argued by many authors to play a large role in categorization and similarity judgment (Medin, Goldstone, & Gentner, 1993, for example). It is believed that context modulates which features people pay attention to when categorizing concepts or judging the similarity between them. Unfortunately, none of the currently available theories of context effects are able to predict which features will be important in a particular context.

The current work has the potential to offer these types of predictions. It seems likely that the features people strongly associate with a concept should be the same features they use in categorization and similarity judgments involving those concepts. Thus, by applying a measure such as collocation, we could determine each feature's importance in these judgments.

Previous work by Zeigenfuse and Lee (2010) supports this idea. As discussed in the previous section on domain differences, here the authors use feature subsets selected using heuristics based two of the measures from this paper, cue validity and collocation, to fit similarity judgments between the animals and artifacts stimuli in the Leuven data set. Echoing our result, they found features selected using collocation best able fit the observed similarities. In some cases, the fit achieved using collocation to be nearly as good as their benchmark measure. This suggests that a good measure of association, taken in conjunction with a method for translating between associations and importances within specific models, will be able to model the effects of context on categorization and similarity judgments.

Basic-Level Categories

Our results suggest that people do not make judgments of feature importance according to a cue validity. This empirically supports Murphy's (1982) theoretical argument that aggregating the cue validities of all of the features associated with a category is not a good measure of how basic a category is. The argument is that given nested categories, aggregated cue validity will always be higher for the more inclusive categories. For example, the category *animals* contains the category *bird*, and so has the higher cue validity, but empirically, *bird* has been found to be a basic level category but *animals* has been found not to be (Rosch et al., 1976).

Our results support his argument as follows. Suppose a category is determined to be basic on the basis of large cue validities for associated features. If we then turn around and ask people to produce cue validities for features associated with the basic category, the cue validities of these features should be high. As a consequence, if people use cue validity to determine which categories are basic, then cue validity should best predict their feature importance rankings for basic categories. Our set of concepts includes basic categories. Thus, people's preference for features ranked highly by other measures provides evidence against aggregated cue validity being used to determine category basicness.

Conclusion

In this paper, we have empirically compared three measures of association between concepts and features. These measures were cue validity, collocation, and category utility. We found that collocation captured human associations better than cue validity and category utility, and that category utility captured human associations better than cue validity. Additionally, we found strong evidence for individual or domain differences (or both) in which measure best captures people's associations.

Though these results begin to paint a picture of how people form associations between concepts and features, additional work is needed. Future work should investigate the absolute ability of each measure to capture human associations. Additionally, future work should explore the relationship between the strength of a feature's association with a concept and its role in categorization decisions and similarity judgments involving that concept. This will allow us to develop not only good measures of association, but to understand how these measures affect context.

References

Corter, J. E., & Gluck, M. A. (1992). Examining basic categories: Feature predictability and information. *Pyschological Bulletin*, *111*, 291-303.

de Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M., & Voorspoels, W. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, *40*(4), 1030-1048.

Jones, G. V. (1983). Identifying basic categories. *Psychological Bulletin*, *94*, 423-428.

Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*(2), 253-278.

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, *91*(1), 174-177.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 352-382.

Zeigenfuse, M. D., & Lee, M. D. (2010). Heuristics for choosing features to represent stimuli. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (p. 1565-1570). Austin, TX: Cognitive Science Society.