Feature Relations and Feature Salience in Natural Categories

Jonathan R. Rein (jrein@mail.utexas.edu)

Bradley C. Love (brad_love@psy.utexas.edu)

Arthur B. Markman (markman@psy.utexas.edu)

Department of Psychology, 1 University Station A8000
Austin, TX 78712 USA

Abstract

Feature salience is a poorly understood construct in the study of category knowledge. Independent-feature representations are limited in potential explanations, compared to more complex structured representations. We explore the relationship between features' connectivity in a network of relational knowledge and those features' relative salience, along with their potential for inference. Comparing values of individual features, we illustrate a unique relationship between features' salience, inferential potency, and structural connectivity. We argue that these results provide evidence for the explanatory power of structured representation beyond independent-feature representation.

Keywords: concepts and categories, memory, representation

Introduction

Cognitive scientists often invoke salience as an explanatory construct. When responses are biased toward some items or features over others, differential salience provides a fairly simple account. While such explanations may match intuition, they are not theoretically powerful unless one can further explicate why this difference exists.

One area in which salience has a powerful impact is categorization. For example, more salient dimensions have greater initial weights or biases in mathematical models of category learning (e.g. Kruschke, 1992; Nosofsky, Palmeri, & McKinley, 1994). These weights are assigned according to arbitrary intuition or existing similarity data. The determinants of salience are generally left to mystery. We suggest that this explanatory limitation is partially the result of assumptions about the representational form of categories. The majority of published work assumes (explicitly or implicitly) that categories are sets of unstructured, unrelated features. These independent-feature representations are instantiated in two main ways. One is as a list of verbally described features, such as "has feathers", "has wings", and "flies" for the category *bird*. Feature lists are generally obtained for natural categories by querying semantic knowledge of normal participants (e.g. McRae, Cree, Seidenberg, & McNorgan, in press; Rosch & Mervis, 1975; Tversky, 1977). The other type of independent-feature representation is a spatial one, in which features are numerical values in a vector. For instance, the *bird* example above would be represented as [1,1,1]. This is the preferred format for artificial categories and mathematical models, with the features and their corresponding values usually being defined by the experimenter (e.g. Love, Medin, & Gureckis, 2004; Medin & Shaffer, 1978; Nosofsky, 1986; see McRae, de Sa, & Seidenberg, 1997 for a model employing spatial representations derived from feature lists). Features and dimensions may be differentially weighted in these formats, but weights are generally not influenced by the relationships among features.

In the next section, we describe ways that features' salience can differ when those features are represented strictly independently. We then review previous research that has employed structured representations in categorization. From these findings, we suggest a relationship between features' salience and connectivity in a structure of feature relations, which will motivate the current study.

Independent Features and Salience

There is a limited number of ways in which one can explain salience differences among features, given an independent-feature category representation. Tversky first suggested some of these in his seminal work on features and similarity (1977). One determinant of salience is our natural predisposition toward intensity, a contingent result of the design of our cognitive systems. This is most obviously the case for perceptual dimensions of stimuli, such as bright lights and loud sounds. Salience may also be influenced by the statistics of the environment. Two statistical properties of features that could have a straightforward relationship with salience are diagnosticity and frequency. A common measure of diagnosticity in categorization is cue validity, the conditional probability of category membership given a particular feature, or P(*bird*|"feathers"). A common measure of frequency is category validity, the conditional probability of the presence of a feature given category membership, or P("feathers"|*bird*). One can obtain these statistics through simple arithmetic on tallies of independent features. Critically, both of these measures only involve the relationship between the feature and category label. More complex statistics could incorporate information across features and dimensions.

Features may also differ in salience according to the

context in which they occur. Some features of a category or item will become more or less salient depending on the comparison set of stimuli (Goldstone, Medin, & Gentner, 1993; Medin, Goldstone, & Markman, 1995; Tversky, 1977). Features may also be differentially affected depending on the nature of the task. For example, when people must classify similar stimuli into two categories, the features that distinguish between the two categories become more salient (Goldstone, 1996, Markman & Ross, 2003). In modeling terms, they receive greater attentional weight (Kruschke, 1992, Nosofsky, 1986). Attention is generally not clearly defined, but it is often treated as learned salience. Attention may be allocated across multiple dimensions when none are sufficient for classification but they are jointly predictive. Although this acquired salience involves multiple features, it is not dependent on feature relationships per se, but rather the relationship between features and category labels. In this way, these diagnostic effects are analogous to the simple statistics discussed earlier.

In general, any task context presents uncertainty about which action is optimal. Information that reliably predicts the best action is most valuable. Salience is a marker of this information value.

Structured Category Representation

Although the assumption of independent feature representation has stimulated valuable research, it nonetheless seems incomplete. Murphy and Medin (1985) suggest that our natural categories are not merely bundles of unrelated features. Rather, features are bound together in a systematic way, according to background knowledge. We have theories of how the world works, and these theories provide constraints on which features exist and cohere together to form categories. More important for this discussion, theories impose a structure onto features. Our domain knowledge binds relevant features through causal relations. We know not only "has wings" **and** "flies" independently, but also "has wings" **causes** "flies" relationally.

Murphy and Medin's (1985) theory account matches our common intuitions about the extent of our category knowledge. However, it does little to define precisely what a theory is and how it might be instantiated structurally. Without such a description it is unclear how relational structure would provide any additional impact on feature salience beyond that already accounted for with independent-feature representations.

Sloman, Love, and Ahn (1998) performed a series of studies to elucidate the notion of theories and their effects on category features. In particular, they were interested in the role of theories in feature centrality—how critical a feature is for the coherence of a concept. One way to measure this centrality is to gauge the effect of removing a feature on the coherence of the resulting incomplete concept. The relative ease with which one can remove a feature while preserving the concept is referred to as "mutability". Sloman, Love, and Ahn obtained participants' subjective mutability ratings for the features of four natural categories, using multiple questions employing the feature removal theme. They also obtained several other measures of feature importance, such as salience. Table 1 contains a representative subset of these questions. All 4 mutability ratings were correlated, while none were correlated with the other constructs.

Sloman, Love, and Ahn also acquired representations of participants' theoretical knowledge of the categories in a structurally tractable way. For each category, they presented participants with all of that category's features on a sheet of paper. Participants then drew arrows between features, indicating relations between those features. Specifically, they were instructed to draw arrows between two features if one "depended on" another. That is "flies"→"has wings" is equivalent to "flies" **depends on** "has wings". Such dependency relations subsume causality relations, but allow for other knowledge as well. Participants could also weight links between features according to the perceived strength of the dependency relation. Averaging across participants, Sloman, Love, and Ahn created matrices of all pairwise dependency relations, representing the theoretical knowledge of a category.

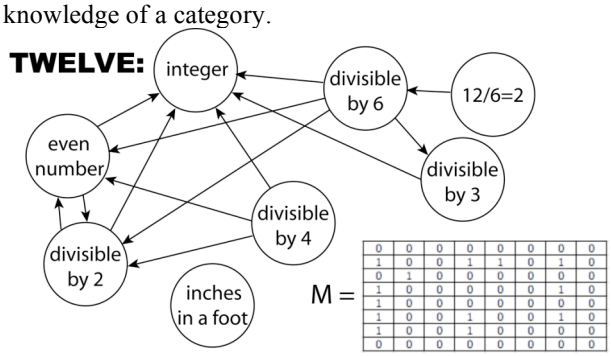

Figure 1: Example dependency graph and matrix

Table 1: Feature Rating Types

Rating type	Example
Salience	How prominent in your conception of bicycles is that they have wheels?
Inferential Potency	If all you know about an object is that it has wheels, what percentage of other features of bicycles would you assume the object had?
Cue validity	Of all things that have wheels, what percentage are bicycles?
Category validity	Of all things that are bicycles, what percentage have wheels?
Mutability	How easily can you imagine a real bicycle that does not have wheels?

Proceeding from this representation, Sloman, Love, and Ahn could structurally define centrality. A relatively simple version derives a feature's centrality from the number of features that depend on it and the strength of those relations. Sloman, Love, and Ahn's formulation also includes the centrality of the dependent features, such that the centrality of a single feature has an iterative, asymmetric quality. They implemented this definition mathematically and illustrated a reliable correlation between structural centrality and mutability ratings.

This structural representation view of categories and theoretical knowledge has proved fruitful in a variety of ways. Ahn and her colleagues have shown that greater feature centrality in a network of dependency relations corresponds with higher typicality ratings, stronger category membership, and other classification indices for exemplars of artificial categories (Ahn, Kim, Lassaline, & Dennis, 2000); differences in the types of "core" features for natural kinds and artifacts in various natural categories (Ahn, 1998); and memory and diagnostic weight for symptoms of clinical disorders by expert clinicians and novices (Kim & Ahn, 2002). Although all of this work is suggestive of a strong role of asymmetric dependency structure in classification, this point is equivocal in light of other evidence.

Rehder and Hastie (2001, 2004; Rehder, 2003) have performed several studies that suggest that dependency per se is not the critical structural element for classification. In these experiments, participants are given background knowledge of artificial categories, some of which have detailed causal relations between features. Importantly, these categories exemplify different causal schemata. In a "chain" network (like that used in Ahn et al. 2000), Feature 1 causes Feature 2, Feature 2 causes Feature 3, and Feature 3 causes Feature 4. In a "common cause" network, Feature 1 causes Features 2, 3, and 4. In a "common effect" network, Features 1, 2, and 3, cause Feature 4. After participants sufficiently learn the category features and their causal relations, they classify transfer exemplars in which the presence/absence of features is systematically varied.

By analyzing the differences in classification according to the presence or absence of a feature (or relation between features), one can determine the classification weight placed on that feature. In the chain and common cause network, the greatest weight is placed on Feature 1, as predicted by the asymmetric dependency model. However, in the common effect network, the greatest weight is placed on Feature 4, the least central feature in terms of dependency.

Structural Connectivity and Salience

The pattern found by Rehder and Hastie (2001, 2004; Rehder, 2003) suggests that classification for these categories is determined less by dependency centrality than overall relational connectivity. On this view, a feature's weight in classification and classification-like tasks (e.g. mutability) is influenced by its connectivity in a structured representation. What about the other aspects of feature importance outlined by Sloman, Love, and Ahn? In particular, does relational connectivity have an impact on feature salience? There are intuitively appealing arguments for such a relationship. First, a feature's salience can be considered roughly coextensive with its level of activation. Assuming activation spreads across features in a manner akin to a semantic network (Collins & Loftus, 1975), the most densely connected features in the network will be those most activated (Steyvers & Tenenbaum, 2005). However activation initially occurs and spreads, elements with more links will benefit.

There is also a more normative argument for the relationship between connectivity and salience. Presumably, salience highlights particular features because they are more relevant or valuable. A pervasive goal of cognitive agents is to obtain more information about their environment. Some features are more valuable to the extent that they promote reliable inferences about other features. Again, the loci of greatest inferential potency are those features that have the most relational links to other features. As with the context effects discussed above, knowledge of these features maximally reduces uncertainty, and is therefore most valuable to an ideally rational agent. Though dependency may not be deterministic, it does express a lawful relation between elements that should aid inference bi-directionally. The boost in salience drawn from connectivity-based inference from relations has been demonstrated previously in the domain of analogy (Clement & Gentner, 1991; Markman, 1997). A similar finding exists in artificial categories for features that are empirically correlated with several other features (Billman & Knutson, 1996).

The current study is designed to demonstrate the existence of this relationship between features' structural connectivity and both salience and inferential potency. To do so, we will obtain values on these and other feature measures, such as cue validity. We will present regression data that supports this posited relationship and disambiguates the relative importance of structural and standard statistical properties. Evidence for such a relationship will suggest that structured representation provides constraints and insights beyond the independent-feature representations traditionally adopted in categorization research.

Method

Participants

118 University of Texas at Austin students participated for partial fulfillment of a course requirement or monetary reimbursement.

Materials

We used eight natural categories: acoustic guitar, apple, chair, robin, caterpillar, cello, eggplant, and saddle. In the interest of consistency, the first four categories are those used by Sloman, Love, and Ahn (1998). These categories have 9, 9, 9, and 14 features, respectively. These features were taken from the feature norms of Rosch, Mervis, Gray, Johnson, & Boyes-Braem (1976). In order to inoculate any findings against any category selection bias, the latter four categories were taken from the feature norms of McRae, Cree, Seidenberg, and McNorgan (in press). These categories have 14. 9, 9, and 9 features, respectively. All eight categories were selected to represent a diversity of natural category types, including both artifacts and natural kinds.

Procedure

Feature Ratings Participants performed a computer task, answering questions presented on the monitor with keyboard responses. For each feature, they answered each of the rating question types presented in Table 1. Questions were presented one at a time. On each trial, the question was displayed over a box that displayed the answer. Participants were instructed to respond to all questions with numerical answers ranging from 0 to 100. One could alter answers until one hit the enter key, after which the trial was over. Following the practice trials in Table 1, all categories and rating types were presented randomly. The task was untimed and self-paced.

Relational Knowledge Assessment The procedure for obtaining participants' theoretical knowledge of feature relations was identical to Sloman, Love, and Ahn (1998). Participants were given a paper packet and three markers. In written instructions, they were informed that they would be reporting their knowledge of relationships between features. Each page in the packet contained the category label at the top, with all of the category's features—individually inscribed in circles—displayed in an oval. Participants were instructed to draw an arrow from one feature to another if the first feature depended on the second. An example of what such a dependency graph might look like was presented on the instruction page, using the number 12 and its mathematical properties. The three markers were blue, green, and red. Participants were instructed to use these markers to indicate low, medium, and high strength dependency relations, respectively. The task was untimed and self-paced.

Results

For each feature, we computed its value on a particular rating by taking the mean response for that feature across participants. Therefore, there is a single data point for each feature, 82 in all.

To quantify feature connectivity, we first converted each participant's dependency graphs into matrices. For instance, each graph for apple was translated into a 9x9 matrix, corresponding to all of the possible pairwise relations that could obtain between features. These are two-dimensional matrices because the dependency relation is asymmetric; participants could and did draw arrows for X **depends on** Y without a complementary arrow for Y **depends on** X. Each cell in the matrix contained a number corresponding to the strength of the relationship: 0, 1, 2, or 3 for none, low, medium, and high, respectively. For each category, we created a summary matrix by taking the mean of each cell across participants. We then obtained connectivity values for each feature by computing the mean relation strength for that feature. For a feature in the apple category, this is obtained by taking the mean of the 8 values in the matrix column and the 8 values in the row (no participant indicated that any feature depended on itself). So, the minimum connectivity value for any feature was 0, corresponding to complete relational isolation. The maximum value was 3, corresponding to high-strength, symmetric dependency relations with all other category features. The mean connectivity value was 0.27.

As predicted, there is a highly reliable positive correlation between connectivity values—defined as mean relation strength—and salience values across all 82 features ($r = 0.47$, $p < .001$). This positive relationship also holds within all eight individual categories (r's = 0.58, 0.54, 0.63, 0.54, 0.71, 0.52, 0.63, and 0.78 for acoustic guitar, apple, chair, robin, caterpillar, cello, eggplant, and saddle, respectively). In fact, all of the correlations for individual categories are stronger than the correlation obtained by collapsing across categories. This suggests that the statistically reliable correlation reported above is not simply an artifact of combining otherwise distinct categories.

Also as predicted, there is a reliable positive correlation between connectivity values and inferential potency values ($r = 0.27$, $p = .02$). This positive relationship holds for six of the eight categories, (r's = 0.42, 0.73, 0.77, -0.25, 0.81, 0.02, 0.71, and 0.81 for acoustic guitar, apple, chair, robin, caterpillar, cello, eggplant, and saddle, respectively). Again, many of the individual category correlations are stronger than the overall correlation. Scatterplots of these data can be found at homepage.psy.utexas.edu/homepage/students/Rein/CogSci07data.pdf.

Although these correlations are reliable, one could argue that some construct other than structural connectivity is influencing both salience and inferential potency. To address this possibility, we performed multiple regression analyses with these values as dependent variables. For each, we used the other four feature importance values (mutability, cue and category validity, salience/inferential potency) and connectivity as predictors. We also used collocation (the product of cue

Table 2: Multiple regression of salience

Predictor	Beta
Category validity	0.65**
Cue validity	-0.01
Collocation	0.08
Inferential potency	0.04
Mutability	-0.17
Dependency	0.08
Connectivity	0.27**

** denotes $p < .01$

Table 3: Multiple regression of inferential potency

Predictor	Beta
Category validity	0.08
Cue validity	0.31*
Collocation	0.48*
Salience	0.05
Mutability	-0.25*
Dependency	-0.20
Connectivity	0.27*

* denotes $p < .05$

and category validity) and mean dependency strength (i.e. just the relations in which the feature was depended on) as predictors. The predictors and their standardized regression coefficients are presented in Tables 2 and 3.

Using these seven variables as predictors of salience, the overall R is 0.86 ($R^2 = 0.74$). The best predictor is category validity, followed by connectivity. All other predictors are not statistically reliable. Predicting inferential potency, the overall R is 0.81 ($R^2 = 0.65$). Collocation, cue validity, connectivity, and mutability are all statistically reliable predictors.

Discussion

As we predicted, features that are more densely connected in a structured relational representation are perceived as more salient. This increased salience may be a byproduct of the greater inferential value of features that are relationally linked with many other features. Our findings suggest that the existence of these two reliable correlations is not merely coincidence. Of all of the measures and transforms discussed, only structural connectivity correlates with both salience and inferential potency. Likewise, the only measures that connectivity reliably correlates with are those two. Amidst all of the shared variance that comes with so many measures of feature importance, this unique relationship is no small thing. Including previous research, connectivity has proven to be a reliable predictor of salience, inference, and classification. This convergence of properties is reminiscent of the basic level phenomenon in categorization.

Of course, structural connectivity does not provide the whole story. Any complex judgment will be multiply determined. For salience and inferential potency, people seem to rely on statistical information along with category structure. Indeed, for each of these, a statistical property was the best predictor, though the exact property differed according to the judgment in question. These properties provide information complementary to connectivity. In the case of salience, the best predictor was category validity, or frequency. A natural way for frequency to increase salience is through better memory retrieval, similar to the benefits of relational connectivity hypothesized above. In the case of inferential potency, the best predictor was collocation, followed by cue validity. It is not surprising that measures of diagnosticity map onto inference; one can infer little from something that is true of everything. In fact, one reason for the negative correlation between inferential potency and connectivity exhibited in the *robin* category is that its highly connected features are non-diagnostic, such as "moves" and "eats". The other strong predictors of inference—dependency and mutability—represent the explicit asymmetry of causality. This affirms the reasonable intuition that people prefer to infer effects from causes rather than vice versa. We can therefore uphold the idea that the direction of causality does play an important role in feature relations (Ahn et al., 2000, Rehder & Kim, 2006).

Simple statistical information clearly has an important role in category knowledge. We have also argued for a strong role for more complex relationships between features. How is this relational knowledge learned? One possibility is that it is gleaned from pairwise correlations between independently-represented features (McRae, de Sa, & Seidenberg, 1997). However, an analysis of these features indicates that many are themselves relational—elements of larger structures containing other categories and features (Jones & Love, 2006). Recent advances in Bayesian modeling have developed techniques for building and selecting these kinds of structured causal models from observable data (Getoor, Friedman, Taskar, & Kollar, 2002; Tenenbaum, Griffiths, & Kemp 2006). Simple statistical information is valuable for its own sake, but statistics can also be used to learn more complex relational structures. The study reported here and other evidence suggest that both sources of information are used in various category judgments (McNorgan, Kotack, Meehan, & McRae, in press; Wisniewski, 1995).

Although others have effectively argued for the importance of structural properties in categorization (e.g. Murphy & Medin, 1985; Rehder & Hastie, 2001; Sloman, Love, & Ahn, 1998), this is the first illustration of such factors' influence on salience and inference. These findings provide an additional explanation of the source of salience, and in so doing provide an additional reason to consider structured representation as a viable and necessary alternative to independent-feature representation. Incorporating structure allows one to preserve the powerful statistical information so commonly used while also introducing relational information that is

often neglected but clearly crucial to several aspects of categorization.

Acknowledgments

We thank Steve Sloman for providing original data from Sloman, Love, & Ahn (1998).

This work was supported by AFOSR grant FA9550-04-1-0226 and NSF CAREER grant #0349101 to B.C. Love.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition, 69,* 135–178.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41,* 361–416.

Billman, D. & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 458-475.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15,* 89-132.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Getoor, L., Friedman, N., Taskar, B., & Koller, D. (2002). Learning probabilistic models of relational structure. *Journal of Machine Learning Research, 3,* 679-707.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*, 608-628.

Jones, M. & Love, B. C. (in press). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology*.

Kim, N. S. & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131,* 451–476

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

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309-332.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592–613.

McNorgan, C., Kotack, R. A., Meehan, D. C., & McRae, K. (in press). Feature-feature causal relations and statistical co-occurrences in object concepts. *Memory & Cognition.*

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (in press). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods, Instrumentation, & Computers*.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*, 99–130.

Medin, D., Goldstone, R., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Medin, D.L., Goldstone, R.L., Markman, A.B. (1995). Comparison and choice: relations between similarity processes and decision processes. *Psychonomic Bulletin and Review, 2,* 1–19.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Murphy, G.L., Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-312.

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

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,*53-79.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27*, 709-748.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91,* 113–153.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

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

Sloman, S., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22,* 189–228.

Steyvers, M. & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41-78.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10,* 309-318.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327-352.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 449-468.