Psychonomic Bulletin & Review
2001, 8 (2), 377-384

Categorization of novel stimuli in well-known natural concepts: A case study

GERT STORMS, PAUL DE BOECK, and WIM RUTS
University of Leuven, Leuven, Belgium

In this study, we investigated to what extent exemplar-based and prototype predictors can be applied to predicting categorization in natural language concepts. Participants categorized novel tropical foods into two well-known natural language concepts: *fruits* and *vegetables*. The results indicate that both the prototype predictors and the exemplar predictors contribute significantly in accounting for the categorization choices but that the contribution of the prototype predictor comes from just a limited number of features.

In a series of influential papers on the cognitive representation of semantic concepts, Rosch (e.g., Rosch & Mervis, 1975; Rosch, Simpson, & Miller, 1976) suggested that artificial category learning is very relevant for studying the mental representation of natural language concepts. Most contemporary authors of review articles (e.g., Komatsu, 1992; Medin, 1989; Medin & Smith, 1984; E. E. Smith & Medin, 1981) still subscribe to the same viewpoint. In a typical category learning experiment (e.g., Medin & Schaffer, 1978; Nosofsky, 1988), participants learn a *pair* of categories (say, *C* and *D*) through the presentation of exemplars of both categories. Once the participants have mastered the correct labels for all the items in the learning set, they are presented with a new set of items (called the *transfer set*) and are asked to classify these new items into one of the learned categories. The two categories are supposed to function as each other's contrast. By manipulating the characteristics of the items in the learning set and in the transfer set, it becomes possible to study the mental representation of the newly learned categories.

In many of these studies, characteristics of learning and transfer items have been manipulated to differentiate predictions of (1) abstract prototype models (e.g., Reed, 1972), (2) rule-based models (e.g., Busemeyer & Myung, 1992; Vandierendonck, 1995), and (3) exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1988, 1992). In most comparisons, exemplar models provided the better predictions (e.g., Estes, 1986; Hintzman, 1986; Medin, Altom, Edelson, & Freko, 1982; Medin, Altom, & Murphy, 1984; Medin & Schaffer, 1978; Medin & Schwanenflugel, 1981; Medin & Smith, 1981; for an overview, see Nosofsky, 1992). Recently, however, evidence has been presented that exemplar-based and rule-based representations may both be guiding categorization choices (Erickson & Kruschke, 1998; Palmeri & Nosofsky, 1995). (For other hybrid models that combine exemplar- and feature-based information, see Ashby, Alfonso-Reese, Turken, & Waldron, 1998, and J. D. Smith & Minda, 1998.)

Generalizing the findings from artificial category learning to learning and using natural language concepts is not obvious, because of the complexity of most natural language concepts. For instance, for most natural language concepts, it is not clear *which features* are important in the categorization decisions, whereas only a few well-defined features are manipulated in most artificial categories.

Nevertheless, in everyday life, because people may be required to make categorization decisions involving natural language concepts, we often come across situations that strongly resemble the categorization experiments whenever new and unfamiliar stimuli have to be sorted into well-learned categories. For instance, new products bought in the supermarket are categorized as "bottles" or "jars" and are labeled accordingly (Malt & Sloman, 1995). In such cases, people use categories that were learned often a long time ago, mostly in childhood. These categorization situations resemble the *transfer phase* of category learning experiments. In these situations, the mental representations of these well-known natural language concepts cannot be studied by manipulating the characteristics of a learning set of items.

In this paper, we explored to what extent exemplar-based and prototype models can be applied to categorization in natural language concepts. An experiment was performed in which different sorts of predictor variables were used to predict categorization of novel food items in well-known natural language concepts (i.e., *fruits* and *vegetables*). Exemplar predictors were based on the summed (rated) similarity to well-known exemplars of these cate-

This research project was supported by Grant G.0353.99 from the Belgian National Science Foundation (Fundamental Human Sciences) to G.S. and P.D.B. We thank Sarah Casaer for assistance in purchasing the stimulus foods and Els Myncke for her help in running the experiment. All the data described in this manuscript can be obtained from the first author (in Excel files) upon simple request. Correspondence concerning this article should be addressed to G. Storms, Psychology Department, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: gert.storms@psy.kuleuven.ac.be).

 Copyright 2001 Psychonomic Society, Inc.

gories, and prototype predictors were based on the extent to which features of both categories applied to the presented items. We investigated (1) whether exemplar-based predictors account better for categorization decisions than do prototype predictors, as is usually found in category learning experiments (Nosofsky, 1992), and (2) whether (following Erickson & Kruschke, 1998, and Palmeri & Nosofsky, 1995) there are individual features that contribute significantly in the prediction, over and above what can be explained by the exemplar-based predictors.

THE EXPERIMENT

To construct a categorization situation that was as close to the transfer phase of category learning experiments as possible, we wanted to select two natural language categories that are presumed to function as each other's contrast category and that are embedded in the same concept hierarchy and at the same hierarchical level. Furthermore, we wanted to choose two concepts for which stimuli could be gathered that were novel to the participants and that could easily be believed to belong to one or the other category. On the basis of these criteria, the concepts *fruit* and *vegetables* were chosen for the categorization task. (In the Method section, we will elaborate on the techniques used to select the contrast categories.) Intuitively, *fruits* and *vegetables* seem to function as each other's contrast category and to have the same level of abstraction (residing under a more abstract category consisting of edible natural foods). Furthermore and equally important, there exists a rich variety of exotic food with which our potential participants would be unfamiliar and that could be used as stimuli to be categorized in either one of these concepts.

Few attempts have been made to apply exemplar models to natural language concepts (for rare exceptions, see Heit & Barsalou, 1996, and Storms, De Boeck, & Ruts, 2000). Before describing the details of the experiment, we comment on the difficulties in applying exemplar models to predict category decisions for natural language concepts and on how predictions for categorization choices can be derived from particular versions of the exemplar theory. In the following, we will concentrate on *superordinate level* natural language concepts (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), because this is the level of abstraction of the concepts *fruits* and *vegetables* used in our experiment. We will also briefly review Hampton's (1979) procedure to derive prototype predictions for natural language concepts.

Exemplar Models

In the context of natural language concepts, it is very hard to derive predictions from an exemplar model with specific memory traces, because it is impossible to find out how many and which exemplar experiences people have stored for concepts at the abstraction level of, for instance, *bird* or *sports*. Assuming that the cognitive system only activates a *sample* of the exemplars stored in memory, a possible sampling process is related to the instantiation principle (De Wilde, Vanoverberghe, Storms, & De Boeck, in press; Heit & Barsalou, 1996; Storms et al., 2000). According to the instantiation principle, a representation of a category includes detailed information about its diverse range of instances, and people generate instances of a category to base category-related decisions on. For example, *mammals* are assumed to be judged typical for the category *animals* to the extent that its generated instantiations (e.g., *dog*, *cats*, *humans*, etc.) are typical animals (Heit & Barsalou, 1996). Also, a particular bird *x* is assumed to be typical to the extent that it resembles other activated instances of the category *birds* (Storms et al., 2000). An instantiation process can lead us to predict that, when asked to classify a novel stimulus into one of two categories, people will evaluate the similarity of the presented stimulus to activated exemplars of the two categories. Note that this instantiation principle does not specify at what level the instantiation is really represented. In other words, when *mammal* is (among others) instantiated with an instance like *dog*, this may refer to the level of highly specific memory traces or to the level of an abstract summary representation of all the experiences a person has had with dogs. Whereas Heit and Barsalou's model states that only a single instantiation is activated, we will assume that multiple instantiations can be activated and that the exemplars of the two rival concepts that are activated are the same exemplars that are generated in an exemplar generation task (see also Storms et al., 2000).

Prototype Predictions

The prototype theory assumes that people have abstract summary representations of superordinate natural language concepts directly stored in their mental lexicon by means of lists of characteristic features (Komatsu, 1992; Rosch & Mervis, 1975; E. E. Smith & Medin, 1981). Following the procedure described by Hampton (1979), the similarity of a set of items to a prototype of a superordinate concept can be derived by first asking participants to generate features of the concept and by summing the applicability frequencies of these features for each of the items (possibly weighting the sum according to some importance criterion). Hampton successfully predicted typicalities and response times of well-known stimuli in superordinate natural language concepts by using this procedure, but the same prototype predictor can be used to predict categorization of novel stimuli in superordinate concepts.

In the experiment, prototype-based and exemplar-based predictions were compared with each other and combined to predict categorization of novel stimuli in two familiar natural language concepts. The predictive power of individual features (or rules) was also evaluated when combined with the exemplar-based predictor.

Method

Participants. Fifty-six last-year high school students and 40 students of the University of Leuven participated in the experiment. Thirty-six high school students participated in the *reminds-me-of task*, and the remaining 20 high school students participated in the *categorization task*. Twenty university students participated in a *similarity-rating task*, and 20 participated in the *feature applicability judgment task*. All the participants participated voluntarily. The university students were paid for their participation.

Material. The two superordinate natural language concepts used in this experiment were *fruits* and *vegetables*. To verify whether these concepts function as each other's contrast category, 20 research assistants were given the standard question for deriving contrast sets (Frake, 1961; Malt & Johnson, 1992; Rosch & Mervis, 1975). They were asked to imagine that they have heard a description of an object and that they are trying to guess what the object is. Half of the participants then read that their first guess was a *vegetable*; the other half of the participants read a *fruit* as their first guess. They were to imagine that the first guess was incorrect and were asked to think about the most plausible second guess. All 10 participants that were asked for the contrast category of *fruit* gave "vegetable" as their second guess. Nine out of the 10 remaining participants gave "fruit" as the contrast category for *vegetable*. The 10th participant answered "a plant." These results confirmed our intuition that the two concepts function as each other's contrast category.

A collection of 30 different tropical fruits and vegetables were gathered that were presumed to be novel to our student participants. These fruits and vegetables were purchased in specialized shops that import food from Central African and Southeast Asian countries. Appendix A lists the names and the continent of origin of the stimuli.

Procedure. The 30 stimuli were presented on plates. All 30 plates were placed in a large room on different tables that were separated from each other by wooden partitions. The participants in all the task groups toured the 30 tables in a fixed direction to complete their task. Their starting table was determined randomly. They were allowed to touch and smell the foods, but they were not allowed to squeeze or taste. The presented stimuli were not cut up. They were first asked whether they knew the presented item, and if so, they were asked to write down the name of the item. After answering this question, the procedure was different for the different task groups.

In the *categorization task group*, the participants were asked to categorize each of the items into one of the two categories, *fruits* or *vegetables*. They were also asked to indicate how sure they felt about their answer on a 10-point rating scale, ranging from 1 (*not sure at all*) to 10 (*very sure*).

In the *similarity-rating task*, the participants rated the similarity of the 30 presented items to eight exemplars each (four exemplars of the concept *fruits* and four exemplars of the concept *vegetables*).[1] These target exemplars for every participant were selected randomly from the eight most frequently generated exemplars of the two concepts (taken from an exemplar generation task described in Storms, De Boeck, Van Mechelen, & Ruts, 1996), with the restriction that similarities to all target exemplars (8 for *fruits*, 8 for *vegetables*) were rated by 10 different participants.[2] (In a study by Storms et al., 2000, the instantiation principle was used to predict typicality ratings and response times for well-known, lexicalized items of eight superordinate natural language concepts, including *fruits* and *vegetables*. They found that the predictive value of the instantiation principle increased as a function of the number of "best" exemplars taken into account, but only up to approximately seven.)

In the *feature applicability judgment task*, the participants were asked to judge whether each feature of a list of 17 features applied to the presented item, by marking (for every item) each feature with a "1" or a "0." These features (presented in Appendix B) were selected on the basis of generation frequency in a feature generation task, carried out by 30 other first-year psychology students. The 10 most frequently generated features for both concepts were selected for the feature applicability judgment task, but 3 features were selected for both *fruit* and *vegetable*, resulting in a set of 17 different features.

In a last task, the *reminds-me-of task*, the participants were asked what the presented item reminded them of. The goal of this task was to find out which *fruit* and *vegetable* exemplars, besides the most generated exemplars from an exemplar generation task, might be important for the categorization task. No indication was given as to the level of abstraction of the answers. Thus, all answers were registered, whether they were very general (e.g., "a fruit"), very specific (e.g., "a granny smith apple"), or in between (e.g., "an apple"). The instructions encouraged them to give more than one answer for every item.

Results and Discussion

The few "yes" answers given by the participants to the question whether they knew the presented item were verified. If the answer was correct (which was very rarely the case), the similarity rating (for participants in the similarity-rating task), the classification decision and certainty rating (for participants in the categorization task), the reminds-me-of answer (for participants in the reminds-me-of task), or the applicability decision (for participants in the feature applicability judgment task) for the corresponding item was discarded from the data set.

Two dependent variables and two predictor variables were constructed. The first dependent variable was the *proportion of "fruit" classification decisions* (which is, of course, in this forced-choice situation, the complement of the proportion of "vegetable" decisions). It is important to note that the proportion of "fruit" classifications differed considerably over the 30 items that were presented. For 5 items, a unanimous classification was given by all 20 participants (one as "fruit," four others as "vegetables"). Furthermore, for 9 of the 30 items, fewer than two thirds of the participants gave the majority classification answer. These results show that the participants differed in their classification choices and thus that the task was not trivial. The reliability of the classification decisions was estimated by applying the Spearman–Brown formula to the split-half correlation, after randomly dividing the participant group into two groups of equal size. The reliability estimate of the classification decisions was .92.

The second dependent variable was also based on the classification decisions but incorporated the *certainty ratings*. All the classification decisions were transformed to certainty ratings that were defined on a 20-point rating scale. These certainty ratings ranged from 1 to 20, where 20 referred to a *"fruit" classification with a maximum certainty* and 1 referred to a *"vegetable" classification with a maximum certainty*. "Vegetable" and "fruit" classifications with certainty ratings of 4 were thus transformed into certainty ratings of 7 and 14, respectively. The reliability of the certainty ratings was again estimated using the procedure described above. An estimation of .98 was obtained.

The two exemplar-based predictions (one for *fruit* and one for *vegetable*) were derived as follows. The *instantia-*

tion predictions for *fruits* and for *vegetables* were calculated by simply summing the similarity ratings over the eight most frequently generated exemplars of the corresponding category. Different weightings were tried out, based on generation frequencies and rank order information in the exemplar generation task. Since none of these weightings improved the prediction, only analyses based on the unweighted sum are reported here.

To derive the two prototype predictions, Hampton's (1979) procedure was followed. Two prototype scores were calculated. For every item, the applicability frequency of the item to all the 10 *vegetable* features were summed, and the same was done for all the 10 *fruit* features. The feature frequencies were weighted on the basis of generation frequency of the corresponding features. (Several feature weightings were tried out, based on the generation frequency of the features and on independently gathered feature importance ratings. The weighting based on the generation frequencies yielded the best predictions. Therefore, in the remainder of this article, we will present only the results of this weighted prototype measure.)

Regression analyses were done, in which the two dependent variables were predicted from the two prototype-based predictors (for *fruits* and for *vegetables*) and from the two exemplar predictors, respectively. The prototype predictors accounted for 72.9% of the variance in the categorization decisions and for 71.1% of the variance in the certainty ratings. Both prototype predictors contributed significantly to the prediction of the two dependent variables ($p < .01$). Note that the correlation between both prototype predictors was $-.45$ ($p < .05$). The exemplar predictors accounted for a slightly larger proportion of the variance of both dependent variables: 76% and 77% for the categorization decisions and the certainty ratings, respectively. Again, the exemplar predictors of both *fruits* and *vegetables* contributed significantly ($p < .01$). The correlation between both exemplar predictors was $-.48$ ($p < .01$).[3]

The results reported above show that the two exemplar predictors yielded better predictions in a multiple regression than did the two prototype predictors. Although the prototype and exemplar predictors were correlated, which is not surprising, the proportion of variance that they had in common differed considerably for both concepts (60% for *fruits* and 26% for *vegetables*). Analyses of variance (ANOVAs) were carried out on both dependent variables to evaluate whether the difference in predictive power was significant. Two variables were defined: (1) the absolute difference of the *z*-transformed categorization frequencies and the *z*-transformed prototype prediction of the corresponding items and (2) the absolute difference of the *z*-transformed categorization frequencies and the *z*-transformed exemplar prediction of the corresponding items. In an ANOVA with the stimuli as cases and the type of predictor as a within-cases variable, it was found that the difference in prediction of the prototype and the exemplar-based variable did not reach significance. A similar analysis with *z*-transformed certainty ratings instead of *z*-transformed categorization frequencies showed that the difference between both sorts of predictor variables for the second dependent variable was not significant either.

Next, the dependent variables were predicted from four predictor variables (two prototype and two exemplar predictors). In these analyses, 86.9% and 86.7% of the variance of the two dependent variables could be explained, respectively. The gain in percentage of explained variance was significant for both dependent variables. For the categorization frequencies, the two prototype and the exemplar predictor for *fruits* contributed significantly ($p < .05$). The exemplar-based predictor for *vegetables* was marginally significant ($p < .07$). For the certainty ratings, all four predictor variables reached significance ($p < .05$).

The analysis described above showed that there is information in the prototype predictors that is not contained in the exemplar-based predictors. There are several possible reasons why the prototype predictors contribute significantly over and above what can be explained by the exemplar-based predictors. First, it is possible that different participants use different strategies, with some participants favoring an exemplar strategy and others preferring a prototype strategy. Second, the exemplar set taken into account to calculate the exemplar-based predictor might be too limited. Third, there might be information in the prototype predictors that is simply not manifested in the direct similarity ratings on which the exemplar-based predictor is based. We investigated these three possibilities in detail.

To evaluate possible interindividual differences in decision strategies, intraindividual correlations were calculated between each of the four predictor variables and the certainty ratings for each of the 20 participants separately. For 14 of these participants, the correlations of all four predictor variables reached significance. For the other 5 participants, the certainty ratings correlated significantly with three out of the four predictor variables. Finally, for 1 participant, two correlations did not reach significance. In conclusion, one can say that the substantial contribution of all four predictor variables is not the result of interindividual differences in categorization strategies, since for most participants the four predictor variables correlate significantly with the categorization data.

A second possible explanation for why both types of predictors contribute is that the exemplar predictors are based on a too limited set of exemplars (eight exemplars for *fruits* and *vegetables*, respectively). To have an idea of how many other exemplars might be important in the categorization decisions, we counted, for every stimulus, the number of *fruit* and *vegetable* answers from the reminds-me-of task that were *not* included in the exemplar predictor (i.e., the number of answers that were not among the eight most frequently given responses in an exemplar generation task). These two vectors of frequencies (one for *fruits* and one for *vegetables*), which will be

called the *exemplar-rest* predictors, were used as predictor variables, together with the two exemplar measures, to predict the two dependent variables. The four predictor variables accounted for 84.7% of the variance in the categorization choices and also for 84.7% of the variance in the certainty ratings. Remarkably, in predicting the categorization choices, the exemplar predictor for *fruits* and the exemplar-rest predictor for *vegetables* contributed significantly ($p < .05$), whereas the contribution of the exemplar predictor for *vegetables* was only marginally significant ($p = .08$) and the exemplar-rest predictor for *fruits* was not significant. For the prediction of the certainty ratings, all the predictors but the exemplar-rest predictor for *fruits* contributed significantly ($p < .05$).

Since these results suggest that more than eight exemplars of the *fruit* and *vegetable* categories might have been activated, we decided to extend the exemplar predictors. The most frequently given answers for every item in the reminds-me-of task that were not also among the items included in the original exemplar predictors were selected to extend the exemplar predictors. Ten first-year psychology students (who were paid for their participation) rated the similarity of each of the stimuli to the items of this extended set. Thus, in the additional similarity-rating task, similarities were rated to six more *fruits* and to five more *vegetables*. There was, however, a difference in procedure: In this additional similarity-rating task, photographs of each of the stimuli were presented (on a background that indicated their size), instead of the actual stimuli themselves. The original fruit and vegetable stimuli were, of course, no longer available, given the lapse of time between both rating tasks.

New *extended exemplar predictors* were calculated by summing similarity ratings to 14 *fruit* exemplars (8 plus 6) and to 13 *vegetable* exemplars (8 plus 5). A regression analysis based on these new exemplar predictors explained 79.6% of the variance in the choice proportions and also 79.6% in the certainty ratings. The two extended predictors reached significance in both analyses. When combining the extended exemplar predictors with the exemplar-rest predictors (i.e., in analyses with four predictor variables), none of the latter contributed any longer, indicating that the extension operation was successful.

To check whether there is still additional information in the prototype predictors that is not contained in the extended exemplar predictors, the two extended exemplar predictors were combined with the two prototype predictors in a regression. This model explained 90.8% and 90.3% of the variance in the choice proportions and in the certainty ratings, respectively. The increase in R^2, as compared with the prediction based on only the two extended exemplar predictors, was significant for both dependent variables. In both analyses, the *vegetable* prototype predictor failed to reach significance, and also the *fruit* exemplar predictor was only marginally significant ($p = .06$ and .08, respectively). In other words, there still is information contained in the *fruit* prototype predictor that cannot be captured by the exemplar measures, even after extending the exemplar predictors.

In order to find out exactly which components of the *fruit* prototype contribute to the prediction over and above what can be accounted for by the exemplar measures, the residuals of the regression with the two elaborated exemplar predictors were correlated with each of the features that were used to calculate the *fruit* prototype predictor. Three features did correlate significantly with the residuals: "is sweet" ($r = .48$), "is/looks tasty" ($r = .55$), and "can get rotten" ($r = .42$). The features "is sweet" and "is tasty" correlated very highly ($r = .82$). These results led us to do regression analyses with the elaborated exemplar predictors and the features "is sweet" and "can get rotten" as predictor variables. The model explained 90% of the variance in each of the two dependent variables, which is about the same as the model with the elaborated exemplar and the two prototype measures as predictors. In the two analyses, both extended exemplar predictors were significant, the feature "is sweet" was also significant, and the feature "can get rotten" was marginally significant (with p values of .057 and .055 for the categorization choices and the certainty ratings, respectively).

GENERAL DISCUSSION

The experiment described in this paper shows that the results of a classification task of novel stimuli into well-known natural language concepts can be predicted very well with exemplar- and feature-based predictors. Exemplar predictors, similar to the measures used by Storms et al. (2000) and based on rated similarities to the most frequently generated exemplars of the studied categories, were shown to predict the categorization decisions slightly better than did prototype predictors, defined as in Hampton (1979), although the difference in predictive power was not significant. We also showed that the exemplar predictors could be improved by taking into account similarities to exemplars that "look like" (at least one of) the presented stimuli. However, the *fruit* prototype predictor contributed significantly to the prediction over and above what could be predicted on the basis of the improved exemplar predictors. Finally, it was shown that the additional information in the prototype predictors was mainly contained in two features of *fruit*: "is sweet" and "can get rotten." Although some of the *fruit* exemplars that were incorporated in the exemplar predictor of *fruit* are clearly sweet and can undoubtedly get rotten (such as, e.g., kiwi, peach, cherry, grape), apparently this feature did not determine the similarity ratings of the novel items with these fruits enough to make these features superfluous in predicting categorization decisions.

What are the implications of this finding, that feature information combined with exemplar information is necessary to optimize the prediction of categorization decisions, for the most common views on concept representation? First, it is not completely in line with the exemplar

view, since similarity with stored exemplars does not seem to be sufficient in predicting categorization. However, it fits nicely with recent findings of Erickson and Kruschke (1998) and of Palmeri and Nosofsky (1995), where both rule induction and exemplar encoding were needed to account optimally for categorization choices in category learning experiments. Second, our findings do cast some doubt on the importance of prototype-like summary representations in making the categorization choices. Supporters of the prototype view might still argue that the prototype can be characterized with a few important features, rather than with a summary score involving many features, but this conception of a prototype is rather far removed from the generally accepted notion, as presented in, for instance, Hampton (1979), Reed (1972), and Rosch and Mervis (1975). Finally, the findings might even be interpreted as evidence that key features, not rules or prototypes, play a role in categorization.

Recently, Nosofsky and Johansen (2000) have argued that many results presented to support rules-plus-exception models can more parsimoniously be accounted for by a model that assumes only exemplar activation but that uses a flexible similarity notion, based on selective attention phenomena. Given that the similarity measure used in our study comes from direct ratings (with no specific instructions concerning the respect to which similarity had to be judged), the tenability of Nososfsky and Johansen's claim for our data is hard to verify.

It might be considered desirable to replicate this experiment by using two other natural language concepts. However, it is very hard to find two categories that fulfill the necessary requirements for applying the same research paradigm. More specifically, it is difficult to find two categories (1) that can be considered each other's contrast category and that exhaust a well-defined group of entities, (2) for which there exists a large enough set of stimuli that are novel to the participants but that can be categorized in one of the two categories, and (3) for which the better known exemplars of the categories are lexicalized. The fact that such concepts are hard to find does not mean that the described *fruit* and *vegetable* study is unimportant because it deals with quite exceptional categories. It only means that similar categorization situations, which very often occur in everyday life, cannot be studied *using the same procedure*.

REFERENCES

Ashby, G. F., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental Psychology: General*, **121**, 177-194.

De Wilde, E., Vanoverberghe, V., Storms, G., & De Boeck, P. (in press). The instantiation principle re-evaluated. *Memory*.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, **127**, 107-140.

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General*, **115**, 155-175.

Frake, C. (1961). The diagnosis of disease among the Subanun of Mindanao. *American Anthropologist*, **63**, 113-132.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **18**, 441-461.

Heit, E., & Barsalou, L. W. (1996). The instantiation principle in natural language categories. *Memory*, **4**, 413-451.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, **112**, 500-526.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory & Language*, **31**, 195-217.

Malt, B. C., & Sloman, S. A. (1995, November). *Evaluating theories of artifact naming*. Paper presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, **44**, 1469-1481.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 37-50.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 333-352.

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

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 355-368.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **7**, 241-253.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, **35**, 113-138.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 700-708.

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, pp. 149-167). Hillsdale, NJ: Erlbaum.

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

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 548-568.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.

Rosch, E., & Mervis, C. B. (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.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural basis of typicality effects. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 491-502.

Smith, E. E., & Medin, D. M. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar based information in natural language categories. *Journal of Memory & Language*, **42**, 51-73.

Storms, G., De Boeck, P., Van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1-15.

Vandierendonck, A. (1995). A parallel rule activation and rule synthesis model for generalization in category learning. *Psychonomic Bulletin & Review*, **2**, 442-459.

NOTES

1. The participants in the similarity-rating task only judged similarities to 8 well-known fruits and vegetables (4 of each), not to 16 (8 of each). The reason for this was that the task, in which participants were presented with each of the novel items on a plate, would had taken too much time otherwise.

2. In the instantiation process described by Heit and Barsalou (1996), a large set of instantiated exemplars are incorporated in the predictor variable, weighted by instantiation frequency. In our version, only the eight most frequently generated exemplars of the two target concepts were selected. There was a simple practical reason for this limitation: The whole experiment had to be completed in 4 days, because the food items were changing form, color, and smell afterward, owing to rotting. This time limit prevented us from gathering more data.

3. In many categorization models (e.g., Medin & Schaffer, 1978; Nosofsky, 1988), the contribution of the two contrasting categories are not combined linearly, as in the regression analyses described above, but ratios of evidence are used according to Luce's choice rule (i.e., where the ratio of the evidence for one category and the total evidence for all categories in the contrast set is calculated). For our data, these ratios always accounted for a substantially smaller proportion of the variance than did the linear regression models.

APPENDIX A

Name*	Proportion of Fruit Classifications	Exemplar Prediction		Prototype Prediction	
		Fruit	Vegetables	Fruit	Vegetables
Lemon grass	.00	11.6	31.0	158	306
Mangosteen	.70	47.8	22.5	244	208
Kiwano	.90	4.8	16.5	313	145
Bitter melon	.05	19.0	32.6	222	330
Tomarillo	.85	45.1	2.9	412	124
Okra	.05	17.0	29.7	246	396
Turmeric	.00	12.9	23.6	112	237
Thai eggplant	.10	35.6	29.3	193	277
Guava	1.00	51.1	19.2	454	228
Bergamot	.70	45.4	21.5	303	273
Lilac	.10	16.4	22.5	150	253
Jerusalem artichoke	.00	29.4	27.9	186	335
Rambutan	.95	43.3	14.8	363	107
Edos	.95	48.6	18.1	453	225
Kumquat	.65	4.4	23.1	360	186
Patisson	.55	27.3	15.9	176	222
Kan-toon	.20	13.8	28.6	137	363
Cola nut	.40	37.7	19.1	168	121
Big gourd	.35	3.9	25.5	297	302
Pitahaya	.65	31.7	24.8	341	110
Sweet potato	.00	17.6	31.4	145	358
Tiny korella	.70	35.0	21.6	297	217
Jujube	.90	51.3	21.6	360	224
Cherimoya	.65	36.1	23.2	357	130
Ripe tamarind	.40	2.1	19.7	238	158
Young peper	.85	26.1	19.9	279	179
Banana blossom	.10	17.2	33.2	294	183
Taro	.30	3.9	22.7	112	211
Chayote	.50	42.4	22.1	253	238
Safon	.40	4.6	24.0	270	200

*The names of the items here are in different languages. It turned out that it was impossible to find English names for all the items.

(*Continued on next page*)

APPENDIX B

Features of *Fruit*

1. is healthy
2. is juicy
3. rich in vitamins
4. is edible
5. contains seeds
6. grows on trees
7. is sweet
8. is tasty
9. is the fruit of a plant
10. can get rotten

Features of *Vegetables*

1. is healthy
2. is edible
3. rich in vitamins
4. is eaten together with potatoes and meat
5. is a plant
6. is cultivated
7. needs to be cooked
8. grows under or just above the ground
9. grows in the garden
10. is green

(Manuscript received September 29, 1999;
revision accepted for publication June 5, 2000.)