## Title
Varying Abstraction in Categorization: a K-means Approach



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

**ISSN**
1069-7977

**Authors**
Storms, Gert
Verbeemen, Timothy
Verguts, Tom

**Publication Date**
2005

Peer reviewed

# Varying Abstraction in Categorization: a K-means Approach

**Timothy Verbeemen (timothy.verbeemen@psy.kuleuven.ac.be)**
Departement Psychologie, University of Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

**Gert Storms (gert.storms@psy.kuleuven.ac.be)**
Departement Psychologie, University of Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

**Tom Verguts (tom.verguts@ugent.be)**
Vakgroep Experimentele Psychologie, Ghent University, H. Dunantlaan 2
B-9000 Ghent, Belgium

**Abstract**

In this paper we propose, instead of the traditional distinction between prototype and exemplar models, a generic model that assumes a continuum between prototypes and exemplars. The model is based on the very successful GCM and an associated prototype model that both assume a representation on continuous dimensions. Abstractions are obtained by taking for each category the centroids of the clusters as produced by K-means clustering, effectively producing the GCM and the Single-Prototype Model as extreme cases. The model was fit on a set of unknown, to-be–classified fruits and vegetables (Smits et al., 2002). Better fit values were clearly obtained for the intermediate solutions indicating a strategy where people compare the test stimuli to a set of multiple prototypes rather than just one prototype or all stored exemplars.

**Keywords:** prototype; exemplar; categorization; varying abstraction; clustering.

## Introduction

Since the ground breaking work of Rosch and Mervis (1975) in the mid-seventies, the idea has gained ground among researchers that one of the most important aspects that define human categorization decisions is similarity. Contrary to what has been called the classical view, people do not seem to base their decisions as to which stimulus belongs to which category on a predetermined set of singly necessary and jointly sufficient characteristics (Komatsu, 1992). In fact, for many of our everyday concepts it appears quite impossible to even formulate such a definition. Rosch showed that several measures related to categorization were in fact related to the similarity of a particular, to-be-categorized, item with its own category and other related categories. In this view, called the *family resemblance view*, there are two possible interpretations of this notion of similarity. A first way is to assume that a category is simply a relational structure, and that membership of an item is simply determined by the similarity relation towards other members and non-members of that particular category. But it is often assumed that categories in fact provide a certain summary or centroid that is determined by a number of weighted characteristics. As such, a category is not just the sum of its members, and membership is not just defined by the relation to other members and non-members. Rather, a new "object" that does not necessarily correspond to a concrete real-world-object, but is an abstraction *over* previously encountered category members, arises. Categorization processes are then assumed to operate on these centroids, rather than on all possible stored members of the categories in question. This second way of thinking about categorization has in fact been the dominant approach in most research dealing with natural language (e.g., Hampton, 1979; Storms, De Boeck, & Ruts, 2000, 2001).

A second line of research (Medin & Schaffer, 1978; Nosofsky, 1986, 1992) focused more on the formal definition of the categorization process. In this tradition one typically uses artificial stimuli that were created in the lab and that have the obvious advantage of being completely under the experimenter's control. Typically, a limited set of training stimuli belonging to two competing categories is presented until people classify these items sufficiently correctly. Consequently, a set of transfer stimuli is presented whose items have to be classified in one of the earlier trained categories. Finally, rivaling formal models are fit to explain the categorization proportions. These competing models express the distinction that was mentioned earlier. Models that assume no abstraction at all, but see categorization as a process that is based on the similarity towards all items that were previously stored in a category, are contrasted with models that assume one central representation for each category. In this tradition the first kind of models are called exemplar models, the second kind are called prototype models. Whereas research in the tradition of natural language has been interested in the representation of concepts in general, the formal approach focused specifically on the distinction between the two notions of categorization, the exemplar and prototype view (Smith & Minda, 1998, 2000).

## Abstraction and Similarity

In the distinction mentioned above between prototype and exemplar models, the emphasis is on the question whether there is total abstraction or no abstraction at all. Such a distinction may seem plausible in the case of only a limited set of relatively similar stimuli with relatively few characteristics to be recalled, as was most often the case in research contrasting the prototype and the exemplar approach. The idea of a single prototype is plausible because of the relative similarity of the stimuli. The idea of no abstraction is sufficiently plausible because of the small number of stimuli and their simple and obvious structure. This reasoning breaks down, however, when one looks at natural language concepts.

Take for example the concept *fruit*. First, we can ask ourselves whether it is possible to have no abstraction at all. In a traditional laboratory experiment, only a few stimuli are presented, in *exactly* the same way. They constitute the full set of exemplars. In natural language, it is not so clear what an exemplar of fruit is. Take, for instance, an *apple*. If we assume no abstraction at all, then classifying an item as belonging to the concept *fruit* would require us to compare it to, among other stimuli, all real-life apples that we have encountered. If no abstraction occurs then one must compare to every specific instance that was ever encountered. To represent a category such as fruit, that would amount to an enormous amount of instances of its members. It appears implausible, therefore, to assume no abstraction at all.

A second question we can ask ourselves is whether it is possible to have a single abstraction for all category members. Returning to our apples, one could argue that abstraction does take place at a certain level. A granny smith could for example be an abstraction over many encountered instances. Or it could be that an abstraction for "apple" exists based on different types of apples. But when we look at a higher category level, this reasoning should at least feel uncomfortable. What would be for instance the abstract representation of the collection of an apple, a litchi and a banana? It seems hard to think of anything that is not absurd or comical.

In more complex natural categories, therefore, it seems that at least some abstraction would have to take place. The question is how. There seems to be no clear reason why abstraction should only take place at the category level. This is all the more clear in natural language where there are categories at different levels. Why, then, should an abstract representation be based on a predetermined category level or an item name, when it is more plausible to say that both abstraction and categorization are similarity-based? When we assume that such abstraction is similarity based, we are saying at the same time that the chance of actually forming an abstract representation is also a function of similarity. One would therefore expect that the amount of abstractions in any group of stimuli would be determined by their internal similarity. In the context of natural language this translates to the idea of *level of abstraction*. Typically, categories such as *fruit* are called superordinates, categories such as apple are seen as basic level categories, and categories such as Granny Smith apples would be called subordinates. As the relative similarity of stimuli in a category decreases when one goes up a level of abstraction, so should the chance of having one unifying abstraction.

## A Generic Framework

In order to formulate a model that can accommodate the idea of varying abstractions as discussed above, we will first define one of the most successful modeling frameworks that incorporates the exemplar/prototype distinction.

In the generalized context model (GCM; Nosofsky, 1986, 1992), an exemplar model, categorization is assumed to be a function of similarity towards all relevant stored exemplars. The model was formulated as a generalization of the Context Model proposed by Medin and Schaffer (1978) to incorporate stimuli that differ on continuous characteristics rather than binary dimensions. In case (physical) dimensions are unavailable, the GCM fitting procedure starts with a multidimensional scaling procedure (MDS; see, e.g., Takane, Young & De Leeuw, 1977) on proximity measures of all stimulus pairs involved. The coordinates of these stimuli are then used as input for the model. In the case of two categories, $A$ and $B$, the probability that stimulus $x$ is classified in category $A$ is given by:

$$\mathrm{P(A|X)} = \frac{\beta_A \eta_{XA}}{\beta_A \eta_{XA} + (1-\beta_A)\eta_{XB}} \tag{1.1}$$

where $\beta_A$ lies between 0 and 1 and serves as a response bias parameter towards category $A$. The parameters $\eta_{XA}$ and $\eta_{XB}$ denote the similarity measures of stimulus $x$ toward all stored exemplars of category $A$ and $B$, respectively:

$$\eta_{XA} = \sum_{j \in A} \exp - \left[ c \left( \sum_{d=1}^{D} w_d \left| y_{xd} - y_{jd} \right|^r \right)^{1/r} \right]^q \tag{1.2}$$

with $y_{xd}$ and $y_{jd}$ as the coordinates of stimulus $x$ and the $j$-th stored exemplar of category $A$ (or $B$ for $\eta_{XB}$, respectively) on dimension $d$. The weight of the $d$-th dimension is denoted by $w_d$, with all weights restricted to sum to 1. The power metric, determined by the value of $r$, is usually given a value of either 1 or 2, corresponding to city-block and Euclidean distance, respectively. The sensitivity parameter $c$ determines the overall scaling of the distances. The parameter $q$ determines the decay of similarity as a function of distance, where typically the values 1 or 2 are used, corresponding to an exponential or a Gaussian decay function.

Much of the traditional research that was based on artificial stimuli used a very limited set of training stimuli that varied on a set of binary dimensions (or features). It would of course be impossible to average over discrete features, so the prototype was conceived as an ideal example of a category and was granted modal values for that category. One of the advantages of the GCM is exactly that it allows one to derive a similarity structure from similarities between stimuli as obtained from actual human judgments. The subsequent representation in terms of a multidimensional space makes it very easy to translate the idea of a prototype as the central tendency of a category (Rosch & Mervis, 1975) into a formal model. The object created by taking, on each dimension, the average coordinate over all members of the category, is a straightforward definition of the prototype. The similarity function changes to:

$$\eta_{XA} = \exp - \left[ c \left( \sum_{d=1}^{D} w_d \left| y_{xd} - \overline{y_{.d}} \right|^r \right)^{1/r} \right]^q \qquad (1.3)$$

where $\overline{y_{.d}}$ denotes the mean value of all stored members of category $A$ on dimension $k$. We will refer to (1.3) in combination with (1.1) as the Single-Prototype Model.

A number of studies have already been conducted that compared prototype and exemplar models (e.g., Nosofsky, 1992; Nosofsky & Zaki, 2002). In many, the GCM performed better than prototype models (but see also e.g., Smith & Minda, 1998, 2000). Recently, we have also applied formal models to the categorization of *natural language* stimuli (Smits et al., 2002; Storms et al., 2000, 2001; Verbeemen, Storms, & Verguts, 2003, 2004; Verbeemen, Vanoverberghe, Storms, & Ruts, 2001). In these studies too, we found an overall advantage of exemplar models.

## Varying Abstraction

In the above presented models, two extremes can be found. First there is the prototype that is seen as the one unifying centroid for the whole category. The other extreme, the exemplar model, corresponds not necessarily to the idea of no abstraction at all, but rather to the lowest level of abstraction under investigation. In laboratory experiments, the exemplars would of course refer to the presented stimuli, and would be truly the lowest level. In natural language research it would be impossible to actually trace all stimuli at the lowest level, i.e. the actual real-life stimuli that were encountered by people. One will therefore have to establish a lower level of interest that is still feasible to obtain. For obvious practical reasons, the approach that is most often used is to define the exemplar level as being one level lower than the category level.

However, in a category with N stimuli, there is a whole spectrum of intermediate abstractions varying from exactly one abstraction, corresponding to the Single-Prototype Model presented earlier, to N 'abstractions' corresponding to the exemplar model (where there is in fact no abstraction at all save perhaps at the exemplar level). Given the fact, then, that there appears to be no reason why the exemplar model should be contrasted only with a model that assumes abstraction over the set of stored category members, what should be the right approach? Abstraction, as defined in formula (1.3), could in fact also be based on any other partition of the stimulus set. With a partition defined as an exhaustive set of nonoverlapping subsets, where category $J$ is partitioned in $K$ different sets $S_k$ one obtains, using the same reasoning we used in the case of the prototype model, $K$ different centroids for each dimension $d$ of a particular category:

$$\overline{y_{.kd}} = \frac{1}{N_k} \sum_{j \in S_k} y_{jkd} \qquad (1.4)$$

Using the above formula, we can formulate a generic model for (1.2) and (1.3):

$$\eta_{XA} = \sum_{k \in A} \exp - \left[ c \left( \sum_{d=1}^{D} w_d \left| y_{xd} - \overline{y_{.kd}} \right|^r \right)^{1/r} \right]^q \qquad (1.5)$$

It is easy to see that this model incorporates the special cases where $K$=1 and $K$=$N$. In the first case, the partition is made up of all stored category members and hence the mean weights correspond to the Single-Prototype Model as described earlier. In the second case, each partition contains exactly one exemplar, and hence the mean weights on each dimension are the original exemplar weights. We will refer to this model as the Varying Abstraction Model. (See also Vanpaemel, Storms, & Ons, submitted).

## Defining the Partitions using K-means Clustering

The Varying Abstraction Model seeks to determine which partition gives the best fit to the data, instead of merely comparing the two most extreme cases. The obvious way to do this is to fit the model to the data, with all possible partitions, and to pick out that model which uses the optimal partition with regards to the categorization data. This would be feasible for datasets with only a limited number of training stimuli or supposed stored members, but for large datasets this would become unpractical or even impossible as the number of possible partitions of a set increases drastically with the set size. To give only a small example, the number of partitions for a set of 5 stimuli equals 52. With two categories with five stored members each this would amount to 52x52=2704 models to be fitted, a large but still computationally feasible number. The number of partitions for a set of 10 stimuli already equals 115975 and would amount, in the case of two categories with 10 stored members, to 13450200625 models to be compared. For categories with a large number of stored members, such as natural language categories, a different approach will therefore be required.

We mentioned already that abstraction, if it takes place at all, should be based on the same principles as categorization: similarity. Partitions of a category, and the associated centroids, should therefore be based on the internal similarity of that category. Not only should very similar stimuli be allowed to merge into a single prototype, but very dissimilar stimuli should be allowed to remain separate as a reference object for the to-be-classified items. Such an approach naturally leads us to consider clustering techniques of some sort. Given the fact that cluster centers for each partition in the varying abstraction model are based on the mean values of the stimuli belonging to that partition, an immediate choice would be K-means clustering (see, e.g., Hastie, Tibshirani and Friedman, 2001).

In *K*-means clustering, one seeks to optimally partition a set of *N* items in a predefined number of $S_k$ subsets so as to minimize the criterion

$$\sum_{k=1}^{K} \sum_{j \in S_k}^{N_k} \sum_{d=1}^{D} \left| y_{jd} - \overline{y_{.d}} \right|^2 \tag{1.6}$$

This minimum is reached by first assigning the stimuli randomly to the *K* clusters, and then computing the cluster centers. Consequently, the items are reassigned to the closest cluster center and the cluster means are computed again anymore. This process continues until the assignments do not change. Because *K*-means clustering is based on the Euclidean distance between items on a number of predictor dimensions, it can simply operate on dimensions that are prespecified or obtained through multidimensional scaling techniques. It can be seen from (1.4) and (1.6) that the cluster centers obtained by *K*-means clustering follow the previously mentioned definitions of the (multiple) prototypes. The most straightforward way, then, to incorporate the *K*-means approach into the Varying Abstraction Model is to simply use the coordinates of the cluster centers as returned by *K*-means clustering into the model.

This effectively leaves us with a total number of *N* partitions per category, where *N* is the number of stored category members. In the case of two rivaling categories *A* and *B* this leaves us with a maximum of $N_A$ x $N_B$ models to be evaluated.

## An Illustration of the Model

In this section, we will present an application of the *K*-means Varying Abstraction Model to a natural language dataset consisting of the two superordinate categories *fruit* and *vegetables* taken from Smits et al. (2002). The choice of a natural language set of the superordinate level allows us to optimally test the model as it seems most relevant for categories that posses, intuitively, different subsets of items that are relatively similar within the subsets but rather different between subsets.

Smits et al. (2002) analyzed a stimulus set consisting of pictures of 79 well-known items, retained after an exemplar *generation* task for the categories *fruit* and *vegetables*, and 30 fruits or vegetables, mostly exotic, that were completely unknown to participants. Ten participants completed a feature applicability task for all stimuli, for the 17 most frequently generated features for *fruit* and *vegetables,* generated by a different group of thirty participants. (Taking the most frequently generated features ensures that the analysis is not clouded by potentially unreliable features that are important to only a few subjects.) A similarity matrix was then obtained by correlating the feature applicability vectors for all 109 stimuli, after summing over participants. A different group of thirty participants classified the well-known stimuli as belonging to either *fruit* or *vegetables*. A group of twenty different participants did the same for the novel stimuli.

In order to obtain dimensions, the derived similarities between the 109 old and novel fruits and vegetables were analyzed with ALSCAL (Takane et al., 1977). A three-dimensional ordinal solution was chosen that explained approximately 96 percent of the variance.

Smits et al. then predicted category decisions based on the geometric versions of the GCM and the Single-Prototype Model and found a clear advantage of the GCM over the prototype model.

### Fitting the *K*-means Varying Abstraction Model

In the illustration presented here we fitted the different models of the generic family to the classification data of the 30 novel stimuli. The first step is to apply *K*-means clustering to each category separately in order to find the cluster centers, and hence the prototype coordinates. The well-known items are seen as the stored items, as they were generated from memory by actual subjects. This means that there are 35 stored exemplars in the category *fruit* and 44 stored exemplars in the category *vegetables*. To make the comparison even more feasible, we chose not to examine every possible clustering ranging from 1 cluster to N clusters, but to work in steps of 4. Hence, for each of the two categories, clustering was applied to the well-known stimuli based on the three dimensions as produced by ALSCAL resulting in 1, 5, 9, 13,....,N clusters for each version of the model. Thus we have a total of 10 successive steps for fruit and 12 for vegetables, including the extreme cases where *K*=1 and *K*=*N*. This leaves us with an effective number of 10 x 12 models to be fitted corresponding to the different combinations of the cluster levels.

In the next step, we simply used the coordinates as produced by *K*-means clustering as model input to define the coordinates of the *K* reference objects for each separate model.

Consequently, the models were fitted to the categorization data. As we are dealing with the scaling of response probabilities in two categories, the obvious way is to maximize the Likelihood assuming the binomial distribution[1]. In order to compare the models, we use the

[1] This amounts to maximizing the binomial probability of the data arising under a specific parametrization of the model,

$$p\left(D \middle| M\right) = \prod_j \binom{n_j}{r_j} p_j^{r_j} (1 - p_j)^{n_j - r_j} ,$$

where the index *j* refers to the *j*-th to-be-classified exemplar. The number of trials for each stimulus corresponds to $n_j$. Here, we use the proportion of classifications in the category *fruit* as a dependent

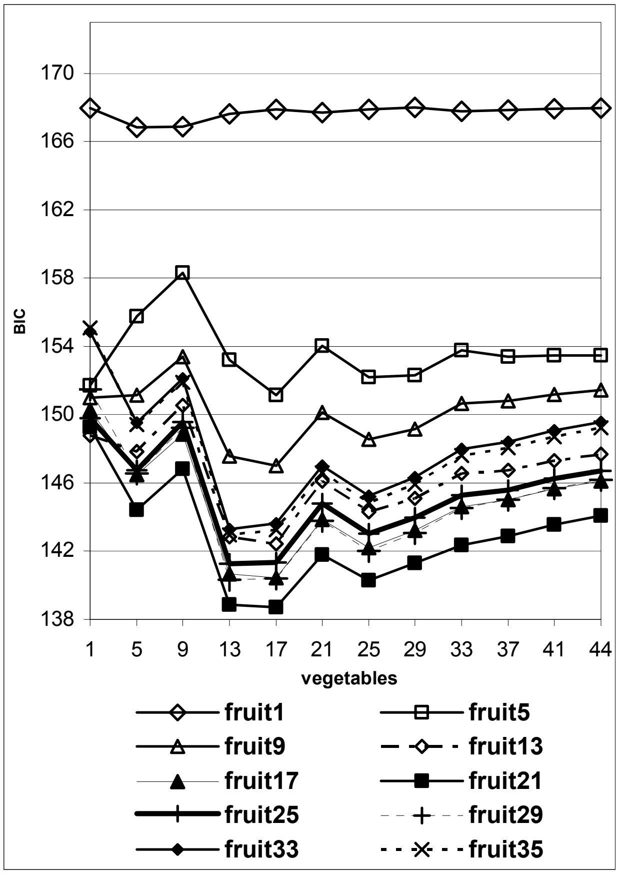


Figure 1a: BIC values for all models. The number of clusters for vegetables is indicated on the X-axis while the different curves represent the cluster levels for fruit.

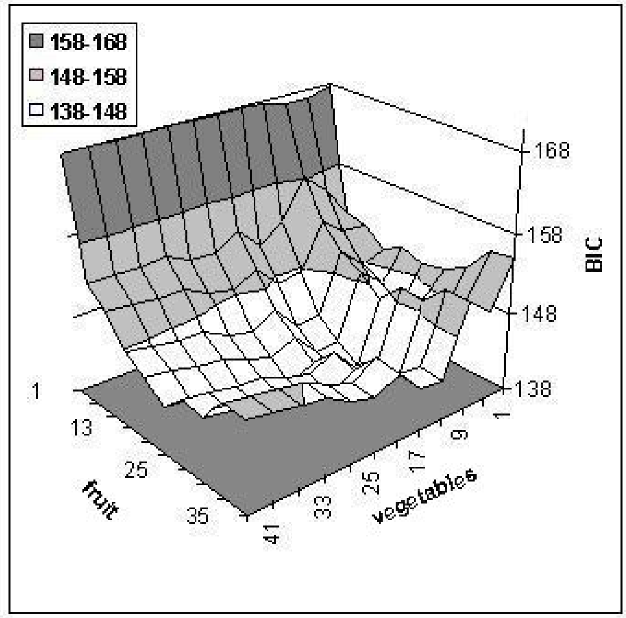


Figure 1b: BIC values for all models as seen on a surface plot. The number of clusters for each category is indicated on the axis.

Bayesian Information Criterion (BIC) that is most suitable for nonnested models as is the case here[2].

**Results and Discussion** We will only discuss models fitted with an exponential decay function (q=1) and Euclidean distances (r=2) as this resulted in clearly better fit values. The results are summarized in Figures 1a and 1b. It is clear from these results that the actual optimum is not situated at the full exemplar model. The BIC value of the exemplar model, which corresponds to the classical GCM, is 149.22. For the Single-Prototype Model, the fit value was 167.97. The model that fits best in our analyses is the model with 21 clusters for fruit and 17 clusters for vegetables. It has a BIC value of 138.71. This difference is large enough to decidedly reject the full exemplar model as the best-fitting model. Furthermore, as can be seen from figure 1a and 1b, there is a relatively smooth decrease towards this minimum, indicating that the optimum is not just due to spurious factors. We can therefore safely assume that the true optimal value is situated at least somewhere around this minimum. Thus, the categorization performance of people seems more likely to be explained by a strategy where people in fact use intermediate abstraction. It does seem to be the case, however, that the models that are relatively more close to the full exemplar model perform better than the models that are relatively close to the Single-Prototype Model.

There seems also to be a certain asymmetry concerning the amount of clusters that are required for the category fruit and the category vegetables. To reach the optimal value, fruit requires more clusters than vegetables even though it has less stored exemplars to be compared with. Furthermore, the fit values as a function of the amount of clusters decrease faster as a function of the number of clusters in fruit than in vegetables. An explanation may by suggested by looking at the actual cluster assignments as produced by *K*-means clustering. In the case of vegetables, there are always less singular exemplars that make up a cluster in any of the cluster levels, except of course for the case where $K=N$. In the case of fruit, a large portion of the

---

variable, so $r_j$ corresponds to the number of trials that the $j$-th item was classified as belonging to fruit. $p_j$ corresponds to the *predicted* proportion of classification of the $j$-th item in the category fruit. Practically, the natural logarithm of the Likelihood function is used as this produces identical parameter estimates.

[2] BIC = -2 ln(L) + k ln(n), where L is the likelihood value, k is the number of free parameters, and n is the number of data points. As such, the measure is a trade-off between model fit and model complexity *Lower* means better, and only the difference in free parameters needs to be taken into account, hence the models presented here are evaluated using only -2 ln(L). The *absolute* difference $|\Delta|$ between two models can be roughly interpreted on a scale of $e^{|\Delta|/2}$ where this approximates the probability ratio of the best fitting model over the worst fitting model (For an extensive discussion, see Kass & Raftery, 1995.)

clusters are singular exemplars, with for instance rhubarb as a single cluster from the start on. In the actual optimum, 14 of the 21 clusters for fruit consist of exactly one exemplar. For vegetables, only 5 of the 17 clusters consist of a singular exemplar. Furthermore, most of the instances of fruit that were kept as singular clusters were highly atypical such as meddler or pomegranate whereas this was much less the case for vegetables. It appears that especially in the case of fruit, the capacity of the model to both keep similar items in one cluster and on the other hand keep outliers separate as a reference object is a crucial feature to explain the categorization data.

## Conclusion

In the present paper, we present a generic model with varying levels of abstraction that incorporates the Generalized Context Model on the lowest level of abstraction, and the Single-Prototype Model on the highest level of abstraction, as special cases. Abstraction is based on similarity and is formally implemented by using *K*-means clustering to find the appropriate partitions within each category. The model therefore allows one to analyze large datasets as it does not require all partitions to be examined.

In an application on a set of unknown to-be-classified fruits and vegetables, the model clearly favors the intermediate levels of abstraction rather than one of the two extremes proposed by the classical models. It seems that especially in the case of larger categories with sufficiently different stimuli, such as natural language categories, this model provides a promising approach to modeling people's categorization decisions.

Finally, the general framework of the model can easily be expanded to other models that use a multidimensional representation. In general, the framework of basing partitions on a similarity-based heuristic could also be expanded to models that use other representational assumptions than the multidimensional models (e.g., Verbeemen et al., 2004).

## Acknowledgments

The first author is a research assistant of the Fund for Scientific Research – Flanders. This project was in part sponsored by grant OT/01/15 of the University of Leuven research council to Gert Storms.

## References

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

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference and prediction*. New York: Springer-Verlag.

Kass, R. E. , & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.

Komatsu, L. K. (1992). Recent views of conceptual Structure. *Psychological Bulletin*, **3**, 500-526.

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

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

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

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **28**, 924-940.

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

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

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **26**, 3-27.

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin and Review*, **9**, 836-844.

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

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of unknown stimuli in well-known natural language concepts: a case study. *Psychonomic Bulletin and Review*, **8**, 377-384.

Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.

Vanpaemel, W., Storms, G. , & Ons, B. (2005). *A varying abstraction model for categorization*. Manuscript submitted for publication.

Verbeemen, T., Storms, G., & Verguts, T. (2003). Determinants of Speeded Categorization in Natural Concepts. *Psychologica Belgica*, **43**, 139-151.

Verbeemen, T., Storms, G., & Verguts, T. (2004). Similarity and taxonomy in categorization. In K. Forbus, D. Gentner & T. Regier, (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pp. 1393-1398. Mahwah, NJ: Erlbaum.

Verbeemen, T., Vanoverberghe, V., Storms, G., & Ruts, W. (2001). The role of contrast categories in natural language concepts. *Journal of Memory and Language*, **44**, 618-643.