*Psychonomic Bulletin & Review*
*2008, 15 (3), 630-637*
*doi: 10.3758/PBR.15.3.630*

# Exemplars and prototypes in natural language concepts: A typicality-based evaluation

**Wouter Voorspoels, Wolf Vanpaemel, and Gert Storms**
*University of Leuven, Leuven, Belgium*

Are natural language categories represented by instances of the category or by a summary representation? We used an exemplar model and a prototype model, both derived within the framework of the generalized context model (Nosofsky, 1984, 1986), to predict typicality ratings for 12 superordinate natural language concepts. The models were fitted to typicality ratings averaged across participants and to the typicality judgments of individual participants. Both analyses yielded results in favor of the exemplar model. These results suggest that higher-level natural language concepts are represented by their subordinate members, rather than by a summary representation.

In the last three decades, two computational models have dominated research on the representation of categories, differing mainly in the assumed degree of abstraction of the category representation. On the one hand, the *prototype view* states that categories are represented by a weighted set of features, abstracted from specific instances of a concept (see, e.g., Hampton, 1993; Rosch, 1975). In this view, the concept *fruit* is a summary representation containing information about what fruit is like, on average. In contrast, the *exemplar view* assumes that a category is represented by instances that have previously been encountered (E. E. Smith & Medin, 1981). The concept *fruit*, in this view, consists of representations of different kinds or instances of fruit a person has experienced.

Both views have been rigorously tested, contrasted, and compared in category learning experiments using artificial stimuli (see, e.g., Medin & Schaffer, 1978; Medin & Smith, 1984; Nosofsky, 1992; Reed, 1972; J. D. Smith & Minda, 1998). In the majority of these studies, the results have favored the exemplar view. This raises the question of whether these results, attained in experimental circumstances and with artificial stimuli, can be readily generalized to the mental representation of natural, everyday categories, such as *professions*, *fruits*, *sports*, or *animals*. For one thing, natural language concepts are most often learned erratically and from many different sources, whereas learning in experimental circumstances consists of an explicit learning phase, in which participants are instructed to meticulously encode the exemplars (Malt & Smith, 1984).

It is quite remarkable that the exemplar view, despite its success in artificial category learning, has been relatively absent in the domain of natural concepts, a domain largely dominated by prototype-like models (e.g., Ameel & Storms, 2006; Hampton, 1979, 1993; Rosch & Mervis, 1975). Only a few attempts have been made to implement the ideas of exemplar theory in models of natural language categories (see De Wilde, Vanoverberghe, Storms, & De Boeck, 2003; Heit & Barsalou, 1996), and comparisons of exemplar and prototype accounts of natural concepts are equally rare (see Smits, Storms, Rosseel, & De Boeck, 2002; Storms, De Boeck, & Ruts, 2000). Consequently, the question of whether natural language categories are represented by a prototype or by different instances of the category remains to be answered more definitely. In this study, we compared a successful exemplar model of categorization, the generalized context model (GCM; Nosofsky, 1984, 1986), with a straightforward prototype model derived from the GCM, within the natural language domain, in order to address this question.

When applying models from the artificial category research to the domain of natural language concepts, two considerations are in place. First, whereas in category learning experiments with artificial stimuli the dependent variable typically is a categorization decision, this variable seems rather awkward for semantic concept research. Adult language users generally are in wide agreement as to whether a certain object or class of objects is a member of language Category A (e.g., *fish*) or Category B (e.g., *birds*). Consequently, categorization decisions show little or no variance across stimuli and participants, and are rather uninformative in the study of the semantic concept representation. Although there are ways to implement category decisions as a dependent variable (Smits et al., 2002), it seems more appropriate to use typicality, or the gradedness within a category: How good an example is an object of a category? Typicality has been an important variable in the study of natural language concepts since Rosch and Mervis's (1975) article on family resemblances and has been shown to be a highly influential variable in a wide variety of cognitive tasks (Hampton, 1993; Malt &

**W. Voorspoels, wouter.voorspoels@psy.kuleuven.be**

*Copyright 2008 Psychonomic Society, Inc.*

Smith, 1984). Typicality has been used to predict performance in tasks of inductive reasoning (Rips, 1975) and production (Hampton & Gardiner, 1983), priming effects (Rosch, 1977), and memory interference effects (Keller & Kellas, 1978).

Second, it is not entirely clear how the term *exemplar* translates to the study of natural language categories. Whereas exemplars in laboratory studies with artificial stimuli can be said to be the visually presented stimuli, a concept like *animal* can be represented by a set of representations of *fish*, *mammals*, *reptiles*, and *birds*, which are in turn abstractions. Alternatively, the concept can consist of the representations of *cows*, *horses*, *dogs*, *sharks*, and *pigeons*, which—again—are abstractions, but are defined at a lower level. This does not mean that the exemplar view is not reconcilable with semantic concepts, since a model does not need to be explicit about how exactly—and on what level of abstraction—exemplar information is stored (see, e.g., Heit & Barsalou, 1996). In this sense, an exemplar-based model can assume a certain degree of abstraction, but only at a level lower than the concept(s) studied. Thus, when studying superordinate language categories (e.g., *fruit* or *vehicle*), members of the category defined at one level lower can be seen as exemplars (e.g., *apple* or *car*). The term *exemplar* in this sense does not refer to an individual, as in its traditional meaning, but to a class at a lower level of abstraction—as in this study.

Taking into account these considerations, we contrasted the GCM with a prototype model derived within the GCM framework with 12 superordinate natural language concepts, using typicality as dependent variable. Although the GCM has already been adapted for predicting typicality (Nosofsky, 1988, 1991), the present study is the first in which the GCM, arguably the most successful model in research on categorization, has been used to predict typicality in natural language categories. It is also the first study in which the GCM and a prototype model have been contrasted using typicalities. The aim of this study is therefore twofold: We want to evaluate the possibility of applying the GCM framework to the study of typicality in natural language categories and, if applicable, to contrast the two traditional approaches to category representation within the natural language domain.

## MODEL REVIEW

### Generalized Context Model

The GCM (Nosofsky, 1984, 1986) essentially assumes that categorization decisions are based on similarity comparisons with individually stored category exemplars. Although the GCM is in essence an exemplar model, one can easily, within its framework, formulate a prototype model (e.g., Nosofsky, 1987, 1992), in which categorization decisions are based on a similarity comparison of an exemplar with the prototypes (i.e., centroids) of the candidate categories.

**Underlying representation**. The GCM assumes an underlying spatial representation in which the exemplars of a category are represented as fixed points in an $M$-dimensional space. The distance between two exemplars in this space is inversely related to their similarity.

**Similarity**. The similarity between two exemplars is calculated from the psychological distance adjusted by dimension weights (reflecting the importance of a certain dimension in the categorization judgments) and a sensitivity parameter, which magnifies or demagnifies the psychological space. Formally, the scaled psychological distance between two exemplars $i$ and $j$ is given by

$$d_{ij} = c\left(\sum_{k=1}^{M} w_k \left|x_{ik} - x_{jk}\right|^r\right)^{1/r},$$

where $x_{ik}$ and $x_{jk}$ are the coordinates of exemplars $i$ and $j$ on dimension $k$, $w_k$ is the attention weight granted to dimension $k$, $M$ is the number of dimensions, and $c$ is the sensitivity parameter. The distance reduces to a Euclidean distance when $r = 2$ and a city-block distance when $r = 1$.

The similarity of a stimulus $i$ to another stimulus $j$ is related to psychological distance, as follows:

$$\eta_{ij} = \exp(-d_{ij}),$$

where $d_{ij}$ is the scaled psychological distance between exemplar $i$ and $j$.

**Typicality prediction**. The exemplar-based measure of the typicality of an exemplar is derived by summing the similarity of that exemplar to all exemplars in the category (Nosofsky, 1988, 1991). Note that, unlike for the prediction of categorization decisions, contrast categories do not affect typicality predictions. Formally, the typicality of an exemplar $i$ for Category A is then given by

$$T_{i\mathrm{A}} = \sum_{j=1}^{n} \eta_{ij},$$

where $\eta_{ij}$ is the similarity of exemplar $i$ to exemplar $j$, with $j$ also belonging to Category A. The free parameters in the model consist of $M - 1$ dimension weights and a scaling parameter $c$.

### Prototype Model

The prototype-based prediction of typicality is derived in much the same way as the exemplar measure, with respect to the underlying representation and similarity, but here typicality of an exemplar in a category equals the similarity of that exemplar to the prototype of the category instead of the summed similarity to all of the category members. The prototype (denoted by $P_{\mathrm{A}}$) of a category is the centroid of the category's exemplars in the $M$-dimensional space (i.e., the point defined by averaging the coordinates of the exemplars on all underlying dimensions). Formally, the typicality of exemplar $i$ for Category A is given by

$$T_{i\mathrm{A}} = \eta_{iP_{\mathrm{A}}},$$

where $\eta_{iP_{\mathrm{A}}}$ is the similarity of exemplar $i$ to the prototype of Category A. The free parameters of the model are identical to those of the exemplar model.

### Fitting the Models

To find the best-fitting parameters for both models, we seeded the parameter space with an initial, randomly chosen parameter configuration and calculated the predicted typicality of the different exemplars of a category given this parameter configuration. We then calculated the correlation between the predicted (i.e., calculated) and observed typicality ratings, fixing $r$ at 2, since it is generally accepted that Euclidean distance is more appropriate for integral dimensions (Shepard, 1964). Next, the parameter space was searched for the values that maximized the correlation. To avoid local maxima, this procedure was used repeatedly and the parameter configuration producing the best correlation was chosen. Overall, the variance among the obtained correlations in the different replications was small, indicating that the solutions were close to optimal ones.

## DATA

For the present study, we used similarity and typicality data for 12 categories, taken from a norm study by Ruts et al. (2004). The set contained 5 animal categories (*birds*, *fish*, *insects, mammals*, and *reptiles*), 2 food categories (*fruits* and *vegetables*), 2 activity categories (*professions* and *sports*), and 3 artifact categories (*musical instruments*, *tools*, and *vehicles*). Every category consisted of from 22 to 30 exemplars, familiar to most adult (Dutch-speaking) Belgians, ranging from exemplars very atypical to very typical of the category. The exemplars of each category were rated by 25 participants for typicality on a 20-point rating scale, with 1 referring to *very atypical or unrelated* and 20 referring to *very typical*. The reliability of the typicality ratings was estimated by split-half correlations corrected with the Spearman–Brown formula, and the values ranged between .90 and .98 (see Table 1 in Ruts et al., 2004). Pairwise similarity judgments were made by a different group of participants on a scale ranging from 1 (*totally dissimilar pairs*) to 20 (*totally similar pairs*). The estimated reliability of the similarity judgments, averaged over 14–17 participants, ranged between .84 and .94 (see Table 1 in Ruts et al., 2004).

## RESULTS

To obtain the representation, the similarity matrices of the 12 categories described earlier (averaged over participants) were used as input for a SAS (version 9) MDS analysis. Since determining the number of relevant psychological dimensions of semantic categories is not a clear-cut task (see, e.g., Lee, 2001; Verheyen, Ameel, & Storms, 2007), representations in 2–7 dimensions were analyzed. The stress values were below .30 for all categories and dimensionalities, and below .15 for all categories from dimensionality 4 onward, decreasing monotonically as a function of dimensionality.

Since it has been argued extensively that averaging across participants alters the structure of the data in a way that might influence the prediction accuracy of competing models (Maddox & Ashby, 1998; J. D. Smith & Minda, 1998), analyses were performed both on averaged and on individual typicality ratings.

### Prediction of Averaged Typicality Ratings

For each concept, typicality ratings—averaged across participants—were predicted for dimensionalities from 2 to 7, using the prototype and exemplar models described earlier.[1] The correlations between the observed and the predicted typicality ratings are presented in Figure 1, as a function of dimensionality for each of the 12 concepts.

Figure 1 shows that the exemplar model's predictions (solid lines) in general were better than the prototype predictions (dotted lines) for all concepts in all dimensionalities from 3 onward, except for *mammals* in 6 and 7 dimensions and *professions* in 3 and 4 dimensions. Although the differences in correlations were not always large, the exemplar model consistently predicted typicality ratings better.

To address this issue further, we conducted a sign test. Looking at the results for 2-dimensional MDS solutions, the exemplar model predicted the typicality ratings better for 9 out of the 12 categories. According to a sign test that assumed a probability of .50 that one model outperformed the other (in other words, under an assumption that both models had an equal chance of "winning"), the probability of observing 9 out of 12 "successes" for the exemplar model was marginally significant ($p = .054$). For higher dimensionalities, the exemplar model predicted typicality better either for 11 out of the 12 categories (for dimensionalities 3, 4, 6, and 7; $p < .01$) or for 12 out of the 12 categories (for dimensionality 5; $p < .001$). Again, although a sign test does not take into account the magnitude of differences in correlation, these results do show that the exemplar model predicted typicality consistently better for different concepts and representations (2–7 dimensions).

It must be noted that in these analyses we assumed that the different exemplars of a category were given equal weight, both in deriving the typicality measure for the GCM and in determining the location of the prototype. This assumption might be problematic, since different exemplars are most probably experienced and encountered with different frequencies. Assuming that generation frequencies in an exemplar generation task reflected the frequencies with which exemplars were encountered, we repeated the analyses after weighting the exemplars of each category with the generation frequency of the exemplars, as reported in Ruts et al. (2004). The results of these analyses were similar to those reported above, indicating that adding exemplar weights to the models did not influence in any major way the basic finding that the exemplar model outperformed the prototype model in predicting typicality.

Although it does not fall within the scope of this study to present a thorough analysis of the reasons for the better performance of the exemplar model, we do want to point the reader to two restrictions of the prototype model, illustrated in Figure 2, that seem to explain the difference in performance, at least partially.[2] First, it seems that the

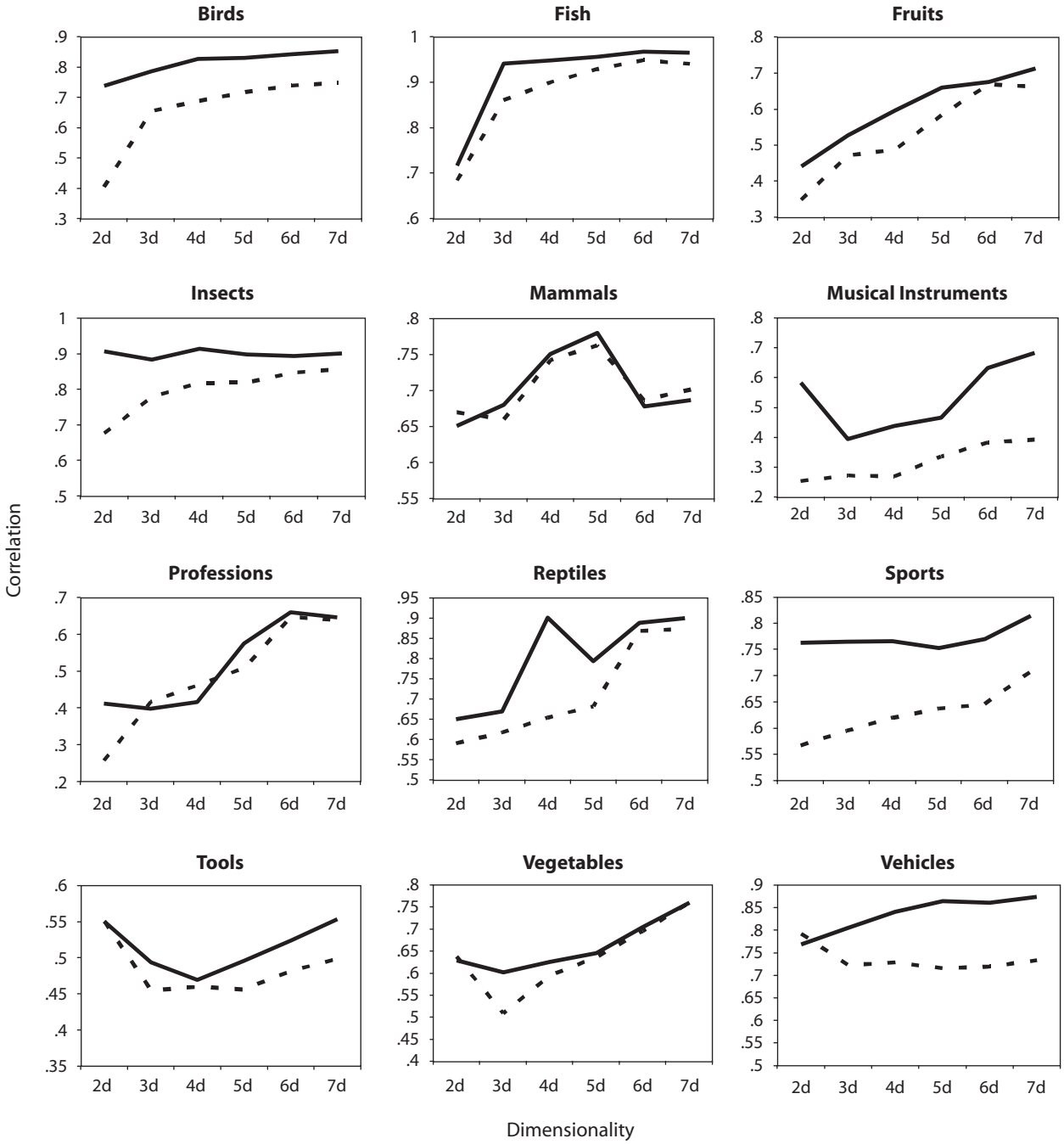


**Figure 1. Correlations between predicted and observed averaged typicality ratings as a function of dimensionality. The predictions of the exemplar model are represented by solid lines and the predictions of the prototype model by dotted lines.**

typicality gradient within a spatial representation of a category does not behave in a symmetrical way, relative to the center (i.e., the prototype). This means that, contrary to what is predicted by the prototype model, exemplars that are symmetrical relative to the prototype do not always have the same observed typicality. Second, the typicality gradient does not always seem to behave unimodally within a spatial representation, since we observed that clusters of relatively typical exemplars can exist at a rather large distance from the prototype.

Figure 2 illustrates these hypotheses, showing, from left to right, a category that was predicted equally well by both models (*vehicles*) and two categories for which the exemplar model clearly outperformed the prototype model (*birds* and *insects*). The restrictions of symmetry and unimodality of the prototype model are obvious in Figure 2. It is also obvious, looking at the categories *birds* and *insects*, that the exemplar model is not restricted in this way, explaining at least partially the difference in the models' performance.

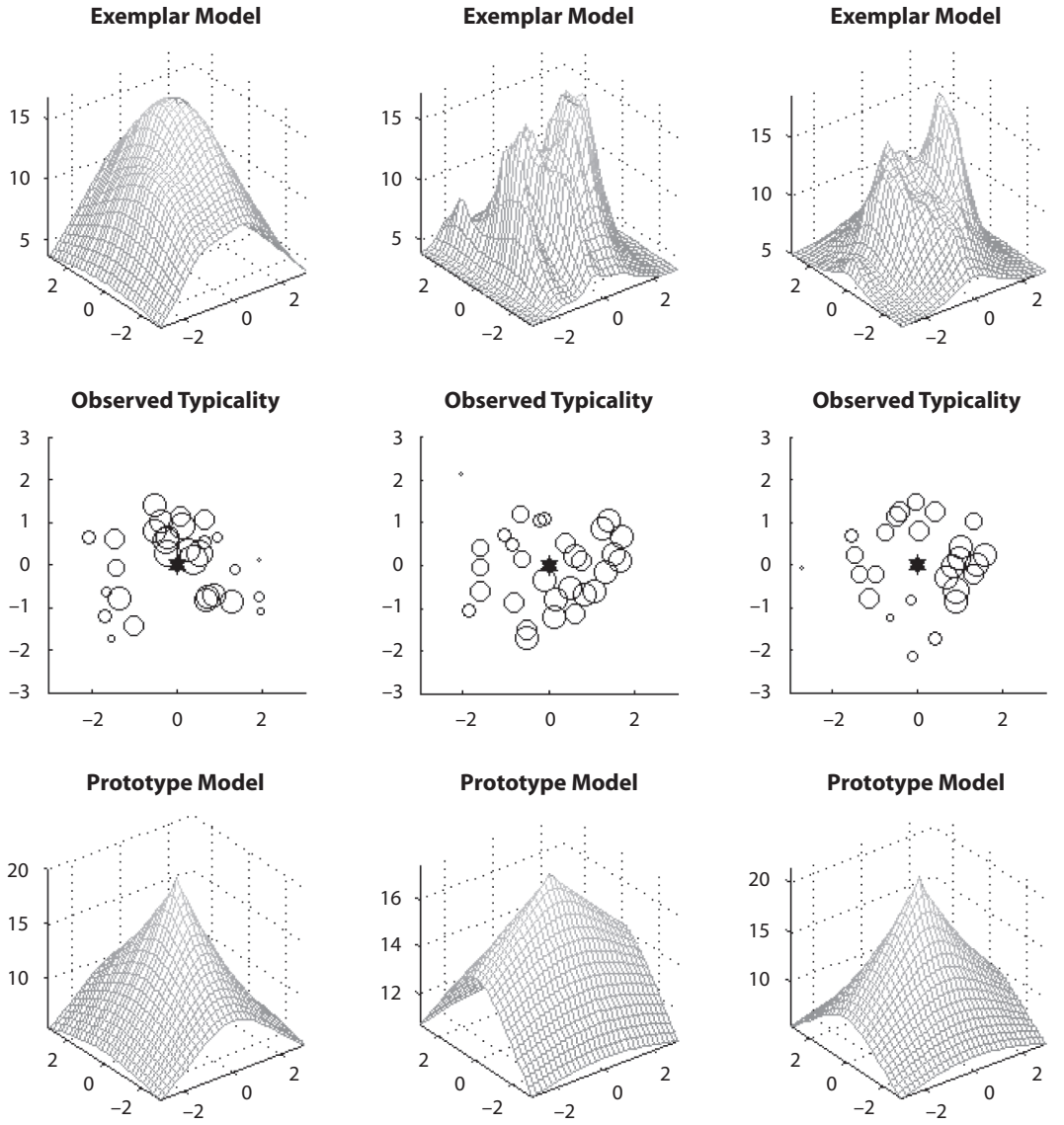


**Figure 2. Observed and predicted typicality for the exemplar and prototype models for the categories *vehicles* (left column), *birds* (middle column), and *insects* (right column). The 3-D plots show the typicality predictions of the exemplar (above) and prototype (below) models for every point in the 2-D representation, calculated with the optimal parameters (including regression parameters—using observed typicality as criterion variable—for reasons of comparability). The plots in the middle row show the exemplars (circles) of a category in a 2-D representational space. The magnitude of the circles refers to their observed typicality, and the black star is the prototype of the category.**

## Prediction of Individual Typicality Judgments

To ascertain that the superior performance of the exemplar model was not due to averaging the typicality ratings across participants, we conducted separate analyses per participant for all categories. Participants showing little or no variance in typicality judgments were excluded from analyses (the criterion used was a standard deviation smaller than 1), since a variance of 0 in typicality judgments leads to an undeterminable correlation with the predictions of any possible parameter configuration, and the optimization algorithm thus could not be applied.

For each concept, both the exemplar and prototype models were fitted to the individual typicality judgments of all participants included in the analyses, again for dimensionalities from 2 to 7. Figure 3 provides an overview per concept of the correlations between the observed and predicted typicality judgments, averaged across participants. Visual inspection of Figure 3 shows that the exemplar model (solid lines) predicted typicality consistently better than the prototype model (dotted lines).

For each category, sign tests were performed for all dimensionalities, under the assumption that both models had an equal chance of predicting an individual's typicality judgments better. Table 1 presents the results of these analyses.

Again, the exemplar model clearly did a better job in predicting typicality: From dimensionality 3 onward, the

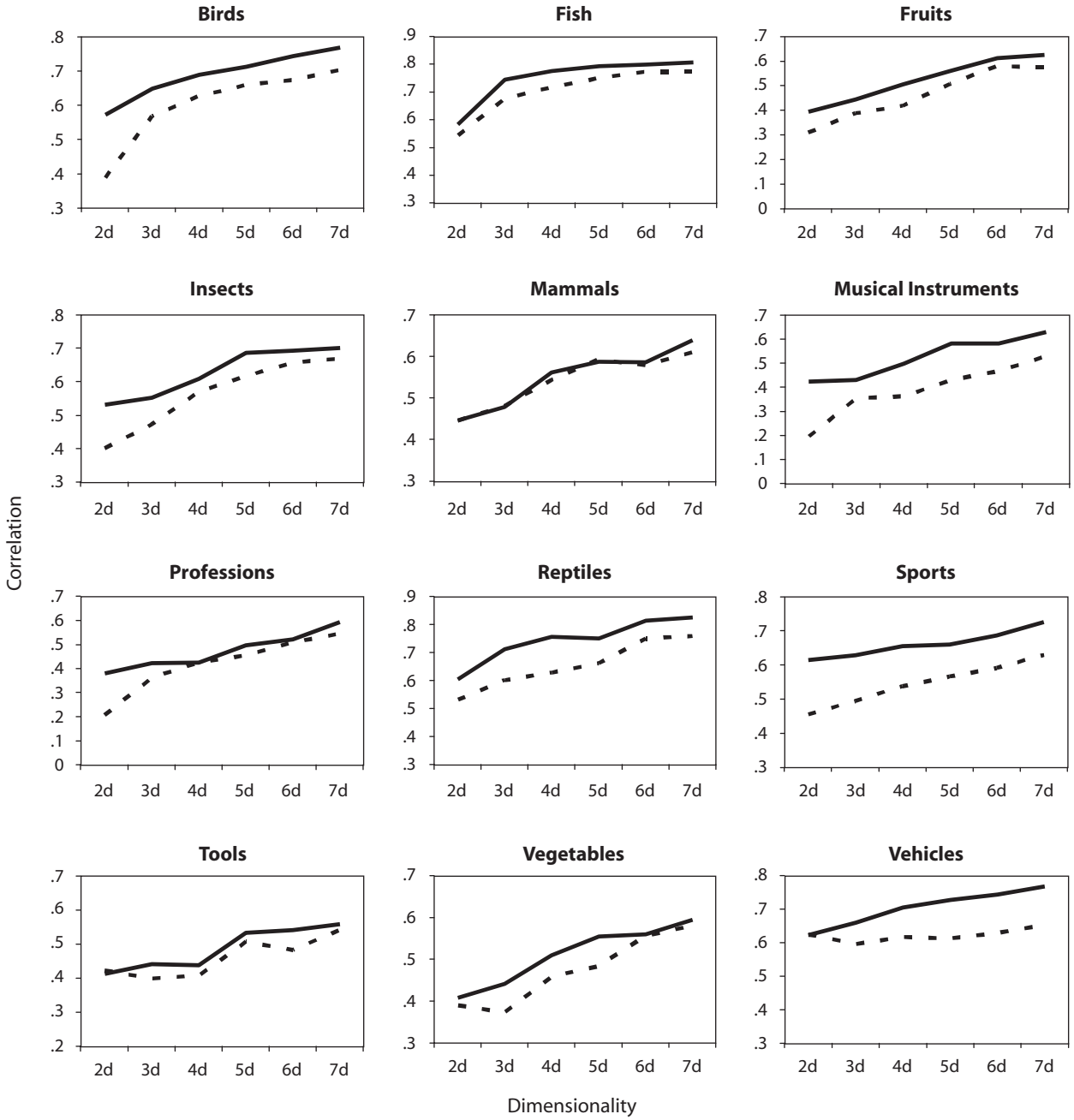


**Figure 3. Correlations, averaged across participants, between predicted and observed typicality judgments per concept, as a function of dimensionality. The predictions of the exemplar model are represented by solid lines and the predictions of the prototype model by dotted lines.**

exemplar model consistently outperformed the prototype model for more participants than one would expect on the basis of the chance level ($p \leq .05$), except for the concept *mammals* in 3, 5, and 6 dimensions; *musical instruments* in 3 dimensions; *professions* in 4 and 6 dimensions; and *tools* in 5 dimensions. Overall, these results are consistent with the analysis of the averaged data, which indicates that the better performance of the exemplar model was not due to averaging the typicality judgments across participants.

## DISCUSSION

In this study, we compared a prototype-based and an exemplar-based account of typicality within natural language superordinate categories. We used prototype and exemplar predictions derived within a version of the GCM framework (Nosofsky, 1984, 1986) that had been adapted to account for typicality (Nosofsky, 1988, 1991) to predict typicality ratings for 12 superordinate natural language concepts. This has been the first study in which typicality

**Table 1**
**Number of Participants for Whom the Exemplar Model Predicted Individual Typicality Judgments Better Than the Prototype Model, per Category and Dimensionality**

| | | Number of Dimensions | | | | | |
|---|---|---|---|---|---|---|---|
| Category | *n* | 2 | 3 | 4 | 5 | 6 | 7 |
| Birds | 23 | 20** | 18** | 20** | 20** | 22** | 22** |
| Fish | 25 | 19** | 23** | 23** | 22** | 21** | 22** |
| Fruits | 25 | 15 | 21** | 20** | 17* | 19** | 18** |
| Insects | 23 | 21** | 22** | 16* | 19** | 20** | 19** |
| Mammals | 17 | 10 | 11 | 13* | 10 | 11 | 12* |
| Musical instruments | 21 | 18** | 14 | 19** | 17** | 21** | 18** |
| Professions | 17 | 15** | 12* | 7 | 12* | 11 | 14** |
| Reptiles | 25 | 20** | 22** | 22** | 23** | 22** | 23** |
| Sports | 25 | 24** | 25** | 25** | 25** | 25** | 24** |
| Tools | 25 | 13 | 20** | 17* | 16 | 21** | 17* |
| Vegetables | 25 | 15 | 21** | 21** | 19** | 18** | 19** |
| Vehicles | 25 | 11 | 21** | 25** | 25** | 24** | 25** |
| Total | 266 | 201** | 230** | 228** | 225** | 235** | 233** |

Note—The sign test assumed a probability of .50 that the exemplar model would predict the ratings better, with *n* referring to the number of participants included in the analyses for a particular category. *$p < .05$. **$p \leq .01$.

judgments for natural language categories were predicted within the framework of the GCM. Also, this is the first time the GCM and a prototype model have been compared using typicality ratings.

Apart from ascertaining that the GCM framework is applicable in modeling typicality in natural language categories, we found that the exemplar model performed consistently better than the prototype model across a broad range of superordinate concepts from different semantic domains (i.e., *animals*, *activities*, *foods*, and *artifacts*), for both averaged and individual typicality ratings. These findings are in line with the main body of evidence in studies of categorization with artificial stimuli, and also with earlier attempts to implement ideas from the exemplar view in the study of natural language categories (e.g., Smits et al., 2002; Storms et al., 2000).

The results suggest that when making judgments concerning the gradedness within superordinate semantic categories such as *sports*, *reptiles*, *fruits*, and *vehicles*, people tend to use information about members of the category (e.g., *soccer*, *snakes*, *apples*, and *cars*) rather than information stored in a prototype.

Two important considerations need to be noted concerning this process. First, our data do not reveal whether including more members of the categories studied would in any major way influence the results. One can imagine, for instance, that people would know more members of the categories *mammals* and *sports* than the ones we included in the stimulus set (which contained from 23 to 30 exemplars). We would, however, expect the contrasted models to suffer equally from this imperfect sampling, which should thus not influence the results in any systematic way.

A second consideration regards the level, if any, on which abstracted summary representations exist. The exemplar model applied in this study does not exclude the notion that abstraction takes place in some natural language categories, since the exemplars we used were natural language concepts themselves, and not exemplars in the strict sense. Although our findings do suggest that superordinate language categories (e.g., *vehicles*) are represented by their subordinate members rather than by an abstracted prototype, this does not imply any representational assumptions concerning these subordinate members; whether or not they are in turn represented by their members is left unresolved.

## AUTHOR NOTE

The research reported in this article is part of Interdisciplinary Research Project IDO/02/004, sponsored by the Research Council of the University of Leuven, as well as of Research Project OT/05/27; both of these grants were given to the third author. Correspondence concerning this article should be addressed to W. Voorspoels, Department of Psychology, ConCat, K.U. Leuven, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: wouter.voorspoels@psy.kuleuven.be; Web site: ppw.kuleuven.be/concat).

## REFERENCES

AMEEL, E., & STORMS, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory & Language*, **55**, 402-421.

DE WILDE, E., VANOVERBERGHE, V., STORMS, G., & DE BOECK, P. (2003). The instantiation principle re-evaluated. *Memory*, **11**, 533-548.

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

HAMPTON, J. [A.] (1993). Prototype models of concept representation. In I. Van Mechelen, J. [A.] Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67-95). London: Academic Press.

HAMPTON, J. A., & GARDINER, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, **74**, 491-516.

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

KELLER, D., & KELLAS, G. (1978). Typicality as a dimension of encoding. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 78-85.

LEE, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, **45**, 149-166.

MADDOX, W. T., & ASHBY, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and

Nosofsky (1996). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 301-321.
MALT, B. C., & SMITH, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning & Verbal Behavior*, **23**, 250-269.
MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.
MEDIN, D. L., & SMITH, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology*, **35**, 113-138.
NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.
NOSOFSKY, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
NOSOFSKY, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87-108.
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. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition*, **19**, 131-150.
NOSOFSKY, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes: Vol. 1. From learning theory to connectionist theory* (pp. 149-167). Hillsdale, NJ: Erlbaum.
REED, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.
RIPS, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.
ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.
ROSCH, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 1, pp. 1-49). London: Academic Press.
ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.
RUTS, W., DE DEYNE, S., AMEEL, E., VANPAEMEL, W., VERBEEMEN, T., & STORMS, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods, Instruments, & Computers*, **36**, 506-515.
SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.
SMITH, E. E., & MEDIN, D. L. (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.
SMITS, T., STORMS, G., ROSSEEL, Y., & DE BOECK, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin & Review*, **9**, 836-844.
STORMS, G., DE BOECK, P., & RUTS, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory & Language*, **42**, 51-73.
VERHEYEN, S., AMEEL, E., & STORMS, G. (2007). Determining the dimensionality in spatial representations of semantic concepts. *Behavior Research Methods*, **39**, 427-438.

## NOTES

1. Note that the Euclidean MDS solutions are rotation invariant, so the "correct" orientation is unknown. Since the estimated parameters stretch and shrink the space along the dimensions of the scaling solution, different orientations would lead to different parameter estimations, and perhaps to different predictions. However, there is no obvious reason why one would expect this rotation invariance to favor either model over the other.

2. To ensure that the better performance of the GCM was not due to greater flexibility, allowing it to fit noisy data better, we fitted both models to data generated from the prototype model, adding different levels of Gaussian noise. The results clearly indicated that, for the generated data, the prototype model outperformed the GCM, suggesting that the better performance of the GCM for the observed data was not due to greater flexibility or functional complexity.

(Manuscript received September 4, 2007;
revision accepted for publication December 21, 2007.)