

# Determining the dimensionality in spatial representations of semantic concepts

STEVEN VERHEYEN, EEF AMEEL, AND GERT STORMS  
*University of Leuven, Leuven, Belgium*

When multidimensional scaling solutions are used to study semantic concepts, the dimensionality of the optimal configuration has to be determined. Several strategies have been proposed to choose the appropriate dimensionality. In the present paper, the traditional dimensionality choice criteria were evaluated and compared to a method based on the prediction of an external criterion. Two studies were conducted in which typicality of an exemplar within a semantic concept was predicted from its distance to the concept centroid. In contrast to the low-dimensional solutions selected by the traditional methods, predictions of an external criterion improved with additional dimensions up till dimensionalities that were much higher than what is common in the literature. This suggests that traditional methods underestimate the richness of semantic concepts as revealed in spatial representations derived from similarity measures.

Over the past 35 years, spatial representations have been used in many models and theories in fundamental and applied cognitive psychology (e.g., Shoben, 1983; Shoben & Ross, 1987). Recently, Gärdenfors (2004) advocated spatial representation as an alternative to the symbol manipulation approach and to connectionism in the study of cognition and illustrated the advantages of spatial representation over the alternatives in domains as diverse as concept formation, inductive reasoning, and the semantics of natural languages. Spatial representations also underlie influential models of categorization, like Nosofsky's (1984, 1986) generalized context model, Ashby and Gott's (1988) boundary decision model, and prototype models (e.g., Medin & Schwanenflugel, 1981; Smith & Minda, 1998).

In most of these models and theories, the underlying representation is embedded in a psychological space, rather than in a space constituted by physical measures of the represented entities. Multidimensional scaling (MDS; Borg & Groenen, 1997; Kruskal, 1964) is a technique that has been of fundamental importance in obtaining the required psychological spatial representation. Typically, researchers gather pairwise similarity (or dissimilarity) data (based on direct similarity ratings, confusion data, sorting data, or response times) as input data. A scaling algorithm then searches for a representation in a given dimensionality in such a way that stimuli are placed closer together the more similar they are and further apart the more dissimilar they are. (For an easily accessible introduction, see Kruskal & Wish, 1978.)

There are applications of MDS in cognitive psychology where the underlying dimensions are perfectly known in advance, and where the contribution of the technique is

limited to psychological scaling of the stimuli along these fixed dimensions. In many experiments in the categorization literature, for instance, stimuli vary along a limited set of well-specified and salient dimensions. These dimensions are also immediately obvious to the participants involved in the task that is used to gather the input (dis)similarities (e.g., McKinley & Nosofsky, 1995). In other applications, however, the purpose of MDS use is twofold: First, researchers want to find out the dimensions that underlie the input (dis)similarities, and second, the stimuli are scaled along these dimensions. In other words, besides psychological scaling, the technique is also used to abstract the underlying stimulus dimensions that constitute the relevant similarity space in which the stimuli are embedded. Often, the resulting spatial configuration is used to predict external variables, like categorization (Smits, Storms, Rosseel, & De Boeck, 2002), response times (Rips, Shoben, & Smith, 1973; Shoben, 1976), linear separability (Ruts, Storms, & Hampton, 2004), inductive strength (Rips, 1975), or typicality (Ameel & Storms, 2006). Moreover, in the neuropsychological literature, patient groups and normal control subjects have been compared to see whether the underlying dimensions of their semantic spaces differ (Storms, Dirikx, Saerens, Verstraeten, & De Deyn, 2003).

When MDS is applied to semantic stimuli to detect an underlying spatial representation, a question that needs to be answered is the number of dimensions of the solution space. Often, subjective criteria are used to determine the dimensionality of the representation and only a brief and shallow account of the chosen dimensionality is given in published reports. In this paper, we focus on the dimensionality of semantic concept spaces. In a first section,

---

G. Storms, gert.storms@psy.kuleuven.be

---

we elaborate on the criteria that have been proposed in the literature to select the appropriate dimensionality of a multidimensional scaling solution. Next, we present a review of published studies where multidimensional scaling solutions were used to study semantic concepts, and focus on the chosen dimensionality and the criteria used to determine the appropriate dimensionality. Then, we present two studies in which spatial representations in different dimensionalities were used to predict an external criterion. Our findings suggest that often informative additional dimensions are rejected when traditional criteria are used to determine the appropriate dimensionality.

### STRATEGIES TO DETERMINE DIMENSIONALITY

Multidimensional scaling algorithms yield, for a certain input data set, the optimal stimulus configurations *in a given dimensionality*. In many applications, however, the appropriate dimensionality is not known in advance. Users therefore calculate configurations in different dimensionalities and then choose one particular solution for interpretation or for further study. Several strategies have been proposed to choose the most appropriate dimensionality. We will briefly elaborate on the most common strategies below.

#### 1. A Priori Theoretical Reasons

Sometimes authors have a priori theoretical reasons to investigate a solution in a particular dimensionality. These reasons are usually very specific for the conducted research.

#### 2. Absolute Stress Values

Often authors evaluate the quality of a MDS solution (in a particular dimensionality) on the basis of the absolute value of the goodness of fit of the solution. The stress of a solution is a fit measure, introduced by Kruskal (1964), that is often minimized in MDS algorithms. The stress equals

$$S = \sqrt{\frac{\sum_{ij} (d_{ij}^* - d_{ij})^2}{\sum_{ij} d_{ij}^2}},$$

where  $d_{ij}$  represents the distance between a pair of stimuli  $i$  and  $j$ ,  $d_{ij}^*$  symbolizes the disparity or optimally transformed dissimilarity for this stimulus pair, and the summation is taken over all possible stimulus pairs of the stimulus set. The denominator is a scaling factor that causes the measure to yield values between 0.0 and 1.0. The optimal transformation is determined using monotonic transformations in nonmetric MDS, and using linear transformations in (the less common) metric MDS. The evaluation of the absolute value of the stress is often based on the guidelines presented by Kruskal (1964). In these guidelines, which were derived from early experience of the author with his own MDS algorithm, stress values of .20, .10, .05, .025, and 0.0 were described as “poor,” “fair,”

“good,” “excellent,” and “perfect,” respectively. However, one should be aware that the stress value depends on the number of stimuli and on the dimensionality. Therefore, Kruskal and Wish (1978) argued that the above described guidelines are only meaningful if the number of dimensions is not too large compared to the number of stimuli that are being scaled. As a rule of thumb, they stated that the number of stimuli should be larger than four times the number of dimensions.

#### 3. Relative Stress Values

Kruskal (1964, p. 16) stated that, when plotting the stress values as a function of the dimensionality, “good data sometimes exhibit a noticeable elbow in the curve, thus pointing to the appropriate (number of underlying dimensions).” This is based on the assumption that the stress decreases systematically with additional “true” dimensions, but that the (slight) decrease due to additional dimensions beyond the elbow just reflects error fitting. Although this reasoning may be appealing in principle, for the vast majority of data sets, no clear elbow can be found.

#### 4. Reliability of the Input (Dis)similarities

A fourth strategy, also already mentioned in Kruskal’s (1964) seminal paper, is that, if information is available about the reliability of the data, one should choose a dimensionality whose stress corresponds to the random component of the data. (See also Borg & Groenen, 1997.) In other words, the (estimated) reliability should equal 1.0 minus the stress value of the solution. A disadvantage of this strategy is that reliability estimates of the input (dis)similarities are seldom available.

#### 5. Interpretability

Interpretability often plays a crucial role in choosing the dimensionality of a MDS analysis. Usually, but not always, the most interpretable solution is chosen from a set of reasonable dimensionalities suggested by goodness of fit. Interpreting dimensions of solutions in dimensionality three or higher may be complicated, often causing authors to stick to only two dimensions. However, regression techniques can be of help to interpret higher dimensional solutions when hypotheses about the underlying features are available (Kruskal & Wish, 1978).

#### 6. Comparing the Stress With Results From Monte Carlo Studies

A more rigorous way to decide on the underlying dimensionality was proposed by Spence and Graef (1974; Spence, 1983) and Wagenaar and Padmos (1971). In this procedure, several data sets are generated, consisting of the same number of stimuli as in the data set under investigation. These Monte Carlo data sets are generated based on a (random) stimulus configuration in a particular dimensionality and with a particular random error component. The underlying (true) dimensionality of these data usually varies from 1 to 4 (Spence & Graef, 1974) or 5 (Wagenaar & Padmos, 1971) and the error level is also manipulated to vary from 0.0 (no error at all) to strongly error perturbed

(e.g., 25% of the variance). All these generated data sets are then analyzed with MDS in different dimensionalities (regardless of the “true” underlying dimensionality, which is, of course, known) and the resulting stress values are stored. Finally, the stress values in different dimensionalities obtained from the empirical data set under investigation are compared with the stress profiles of the generated data to determine the most likely true dimensionality and error level of the data.

### 7. Testing Dimensionality in Maximum Likelihood Scaling

In maximum likelihood MDS models (Ramsay, 1977), the stimulus coordinates in a given dimensionality are estimated by maximizing the probability of finding the empirically obtained data. The model assumes independence of the residuals and a particular error distribution (e.g., in Ramsey’s model: normal or lognormal), thereby turning the model into a metric MDS model. The likelihood of the obtained data, evaluated in the optimally localized stimuli, can then be used to statistically test the optimal dimensionality, using a  $\chi^2$  test, the AIC test (Akaike, 1974), or the BIC test (Schwarz, 1978). (Non-metric maximum likelihood MDS models have been developed, but these models apply only to a limited range of data sets, like seven-point rating scale data. See, e.g., Takane, 1981; Takane & Carroll, 1982.) Although the availability of statistical dimensionality tests is in principle clearly advantageous, this procedure is hardly ever used in the literature because major statistical packages do not include maximum likelihood MDS algorithms. Furthermore, the required test statistics only apply asymptotically (i.e., to very large data sets) and may result in “wrong” dimensionality choices with limited sample sizes (Ramsay, 1980; Storms, 1995).

### APPLICATIONS OF MDS IN THE STUDY OF SEMANTIC CONCEPTS

An elaborate search for MDS applications in the study of semantic concepts (including studies of semantic disorders in the neuropsychological literature) resulted in 43 published papers, most of which described analyses of multiple semantic concepts. In total, 138 analyses were described in these studies. In 3 of these analyses, a solution with one single dimension was chosen (Holyoak & Walker, 1976). In the vast majority (103 analyses), a two-dimensional solution was chosen. Three-dimensional solutions were chosen in 31 analyses, and in only one single study (Arnold, 1971) 4 dimensions were selected. Apart from a study by Ruts, Storms, and Hampton (2004), where solutions in dimensionalities 2 to 5 were investigated without choosing an optimal number of dimensions, solutions with more than 4 dimensions are absent in the literature on semantic concepts known to us. (Furthermore, enlarging the scope of our literature search to other areas of psychology resulted in just a few additional MDS studies where solutions of 5 dimensions were selected.)

For 22 of the 138 analyses, no information regarding the dimensionality selection criterion was mentioned. For

34 other analyses, the only justification was parsimony in the representation, given the limited number of stimuli. More specifically, in all these cases, two-dimensional solutions were preferred because an additional third dimension would violate Kruskal and Wish’s (1978) rule of thumb that the number of stimuli should be larger than four times the number of dimensions. For 26 studies, absolute fit values (stress or the related percentage of variance accounted for) or informal relative fit comparisons (elbow test) were used to select the appropriate dimensionality. For 26 analyses, only interpretability guided the dimensionality choice. Among the latter analyses, there were a few cases in which two-dimensional solutions were selected because people visualize data better in 2 dimensions than in 3 or 4. For 11 other analyses, a combination of fit values and interpretability of the solution was used to select the appropriate dimensionality. A priori theoretical reasons were given for dimensionality selection in 13 analyses. Finally, a mixture of several of the above described selection criteria was used in the remaining six analyses.

Importantly, statistical testing was not used as a basis for selecting the appropriate dimensionality in any of the studies: In none of the studies maximum likelihood models were used, nor was the procedure described by Spence and Graef (1974) or by Wagenaar and Padmos (1971) ever used to check the number of dimensions. Likewise, none of the published studies that we investigated referred in any way to the reliability of the input similarities in the dimensionality selection.

### USING SPATIAL REPRESENTATIONS TO PREDICT EXTERNAL VARIABLES

The selection criteria described earlier can be divided into two classes. The first class of criteria is simply based on the subjective interpretation of the solution without using any quantitative information. This class of criteria includes a choice based on the interpretability of the solutions and a priori theoretical reasons for selecting the appropriate dimensionality. The second class of criteria makes use of the fit of the solution, that is, of the extent to which the model can account for the input data. Any selection criterion based on stress values, in an absolute or a relative sense, or on the likelihood principle belongs in this class.

There is, however, a third kind of selection criterion, which is seldom used in the literature, but which can shed a different light on the choice process. In this selection procedure, information from an external variable (i.e., information not derived from the input similarities) is used to select the dimensionality. This criterion has been used in the domain of personality psychology by Rosenberg and Jones (1972; Jones & Rosenberg, 1974). In their studies, 99 physical and psychological traits were scaled in several dimensionalities. An elbow test suggested three underlying dimensions. Then, ratings for the same 99 traits on 12 properties were regressed on the coordinates of these stimuli in different dimensionalities. Rosenberg and Jones found increasing multiple correlations with in-

creasing dimensionalities up till four dimensions. They interpreted this finding as evidence for useful information in the fourth dimension, despite the fact that the elbow test favored a three-dimensional solution.

In the remainder of this article, we will investigate whether traditional dimensionality choice criteria point at the optimal dimensionality when the ultimate goal is the prediction of an external variable. More specifically, we will show that, in cases where the traditional methods of dimensionality selection prefer low-dimensional solutions, predictions of an external criterion sometimes show systematic information value in much higher dimensionalities.

### CENTROID TYPICALITY PREDICTIONS IN SEMANTIC CONCEPTS

Since similarity-based models of concepts and categories have been proposed as alternatives for the classical view of concepts (Komatsu, 1992), these models have dominated the literature on semantic concepts as well as the literature on artificial category learning. The two most important similarity-based models are the prototype and exemplar model.

In prototype models, it is assumed that concepts and categories are mentally represented by an abstract summary representation of the previously encountered exemplars (Hampton, 1979, 1993; Rosch & Mervis, 1975). Advocates of exemplar models have denied the necessity for summary representations and instead assumed that all previously encountered exemplars are stored in memory (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984, 1986). In this view, category-related decisions are based on the activation of the stored memory traces of these exemplars.

Although predictions of the prototype and exemplar views can be differentiated and contrasted in laboratory experiments with artificial categories, after more than two decades of controversy, the debate between advocates of both views still lingers on (e.g., Smith & Minda, 1998, 2000). In most natural language concepts, in which the exemplars of a concept cluster together tightly and are rather well-separated from nonexemplars, prototype and exemplar predictions are nearly identical and come down to a central tendency representation (Barsalou, 1990). In the context of spatial representations of semantic concepts (which have been proven very useful in the study of semantic concepts by, e.g., Shoben, 1983; Shoben & Ross, 1987; Smits, Storms, Rosseel, & De Boeck, 2002), predictions of both kinds of similarity models are based on the centroid of the exemplar set.

Below, we will describe two studies in which typicality of an exemplar within a semantic concept is predicted from its distance to the concept centroid. Such spatial predictions of typicality or graded structure have been proven effective in several studies before (e.g., Reed, 1972; Rips, 1975). However, in these studies, the dimensionality of the most appropriate representation was usually determined in advance, based on interpretability or on the fit of the solution. Contrary to this common practice, we want to in-

vestigate whether typicality prediction varies as a function of the underlying dimensionality and how this variation relates to the dimensionality chosen based on the common choice criteria.

### STUDY 1

In Study 1, we used similarity and typicality data of 12 semantic concepts, taken from Ruts, De Deyne, Ameel, Vanpaemel, Verbeemen, and Storms (2004). The set of 12 semantic concepts consisted of 5 animal categories (*birds*, *fish*, *insects*, *mammals*, and *reptiles*), 3 artifact categories (*musical instruments*, *tools*, and *vehicles*), 2 food categories (*fruits* and *vegetables*), and 2 activity categories (*sports* and *professions*). In the data from Ruts et al. (2004), the stimulus sets consisted of only clear exemplars of the studied categories and 22 to 30 exemplars were selected for each of the semantic categories. The exemplars, which were presented as words, were rated for typicality within the category on a rating scale, ranging from 1 for *a very poor example* to 20 for *a very good example*. Pairwise similarities among all possible exemplar pairs were rated by a different group of participants on a scale varying from 1 (for *totally dissimilar* pairs) to 20 (for *totally similar* pairs). Ratings were averaged across participants and reliabilities were estimated by the split-half correlations corrected with the Spearman-Brown formula. For the typicality ratings these estimates ranged between .90 and .98. For the similarities the reliability estimates ranged between .84 and .94.

### MDS Analysis and Dimensionality Tests

The similarity matrices for the 12 semantic concepts (averaged over the different participants) were first taken as input for PROC MDS, the MDS procedure available in the SAS package (SAS, V8). This procedure estimates a spatial configuration in a given dimensionality, using a nonlinear least squares algorithm. The default settings were used, yielding a nonmetric MDS analysis. For each concept, configurations were obtained in dimensionalities 2 to 7 (7 being the highest dimensionality that should be accepted for a category consisting of 30 exemplars, according to Kruskal and Wish's, 1978, rule of thumb regarding dimensionality choice). Since the SAS MDS procedure provides a measure of stress for each configuration, we can use these analyses to determine the appropriate dimensionality, according to the strategies outlined above. Figure 1 shows, in solid lines, the stress values as a function of the underlying dimensionality.

Four traditional dimensionality choice procedures were applied to the data: absolute stress evaluation, relative stress evaluation, expected fit based on the reliability of the data, and a Monte Carlo-based procedure. Table 1 shows, for each concept and each selection criterion, the withheld dimensionality.

When evaluating the quality of the different MDS solutions based on the absolute stress values, for each of the concepts the configuration in the highest dimensionality, allowed by the Kruskal and Wish (1978) rule of thumb, is withheld (see second column of Table 1). The chosen di-

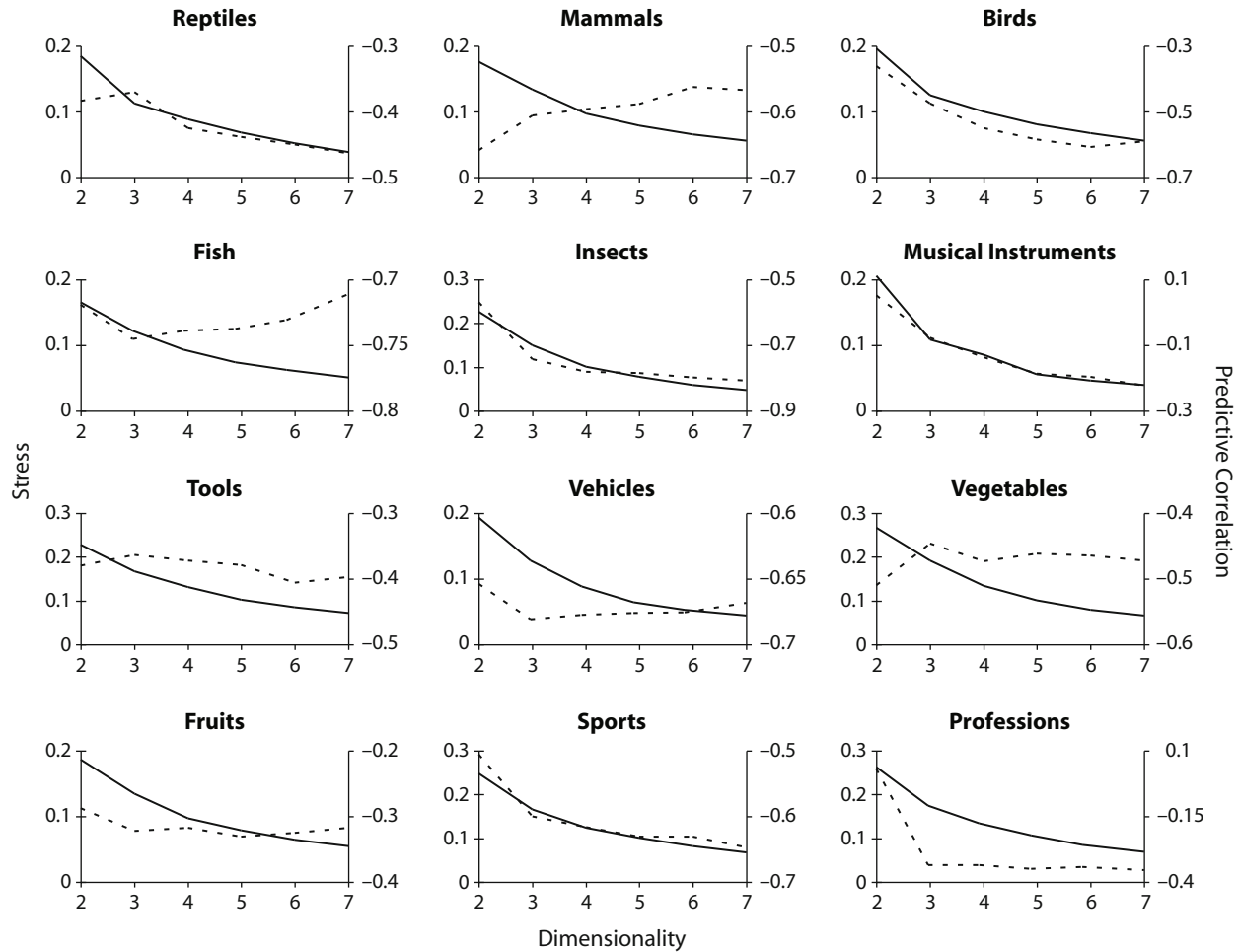


Figure 1. Stress values (solid lines) for MDS solutions in 2 to 7 dimensions for each of the 12 concepts and predictive correlations (dashed lines) for the different concepts as a function of the underlying dimensionality.

dimensionalities varied from 5 to 7. According to Kruskal’s (1964) terminology, all of these configurations would be labeled between “fair” and “good”, except for *vehicles* and *musical instruments*, where the stress value would be described as situated between “good” and “excellent”.

Inspection of the relative stress values (see solid lines, Figure 1) proved difficult since no clear elbow could be found for some concepts. In these ambiguous cases and in the ones where multiple elbows could be found, the configuration in the higher dimensionality was always chosen. This yielded a dimensionality of 3 or 4 for all the concepts except *musical instruments*, where dimensionality 5 was selected (see the third column of Table 1).

The dimensionalities that were determined by choosing the configuration whose stress value best matched 1.0 minus the reliability of the input similarities can be found in the fourth column of Table 1. The resulting dimensionalities varied widely from 2 to 6.

Finally, the dimensionality was also determined based on a Monte Carlo simulation procedure, as described by Spence and Graef (1974) and by Wagenaar and Padmos (1971). Details about this procedure can be found in the

Appendix. For only three concepts more than four underlying dimensions were chosen.

**Centroid Predictions of Typicality**

Often MDS solutions of semantic domains are used to predict cognitive and language-related variables (e.g., Gärdenfors, 2004; Rips, 1975; Smits et al., 2002). One such variable that can straightforwardly be linked to the spatial representation of a semantic domain is typicality (Ameel & Storms, 2006). More specifically, typicality is generally assumed to be related to category centrality. In geometric terms, category centrality is translated as the centroid of the category exemplars.

Centroids were calculated by averaging, for every dimension, the coordinates of the complete exemplar set. Next, Euclidean distances of each of the exemplars to the category centroid were calculated and they were correlated with the corresponding rated typicalities taken from Ruts et al. (2004). This procedure was repeated for the solutions of different dimensionalities. The dimensionality that resulted in the highest correlation between typicality ratings and distances, was selected as the optimal dimen-

**Table 1**  
**Withheld Dimensionality According to Four Traditional Dimensionality**  
**Choice Procedures (Columns 2 to 5) and an External Criterion**  
**(Column 6) for Each of the 12 Concepts in Study 1**

English Category Name	Absolute Stress	Relative Stress	1.0 Minus Reliability	Monte Carlo	Predictive Correlation
Reptiles	5	3	2	3	7
Mammals	7	4	5	2	2
Birds	7	3	4	3	6
Fish	5	3	3	2	3
Insects	6	4	4	4	7
Musical Instruments	6	5	5	3	7
Tools	7	3	4	5	6
Vehicles	7	3	5	5	3
Vegetables	7	4	5	5	2
Fruits	7	4	3	4	5
Sports	7	3	5	4	7
Professions	7	3	6	4	7

sionality. The last column of Table 1 reports the chosen dimensionalities for the different categories.

The dashed lines in Figure 1 display, for every studied concept, the predictive correlation as a function of the underlying dimensionality. Note that the correlations are expected to be negative: the larger the typicality of an exemplar, the smaller its distance toward the centroid is expected to be. Different result patterns are observed for the 12 semantic concepts that were studied (see also the last column of Table 1). For *vegetables* and *mammals*, optimal predictions were obtained with two underlying dimensions. Adding additional dimensions lowered the predictive correlations. For *fish* and *vehicles*, a three-dimensional representation resulted in the best typicality prediction. For *fruits*, *tools*, and *professions*, adding more than three dimensions did not improve, nor worsen, the typicality prediction. For the remaining concepts, however, typicality prediction improved gradually with increasing dimensionality until 6 (*birds*) or sometimes even 7 (*reptiles*, *insects*, *musical instruments*, *sports*) dimensions.<sup>1</sup>

One might raise the objection that the position of the centroid can be influenced by possible outliers. In order to meet this criticism, distances were also computed to a central tendency point defined by the median values on each of the coordinate axes. These distances were correlated with the typicalities in the category. The resulting pattern of correlations between these median-based distances and typicalities did not differ much from the pattern of correlations with the centroid-based distances. For only three concepts (*fish*, *musical instruments*, and *sports*), the median-based correlation, averaged across 6 dimensionalities (2 to 7), exceeded the corresponding centroid-based correlation.

## Discussion

The different procedures of choosing the appropriate dimensionality for the MDS analyses clearly show little consistency. Below, we will describe the findings of the different procedures in detail.

First, the absolute stress value decreases with increasing dimensionality. This tendency is what one expects from a reliable algorithm that does not easily get trapped

in local minima. The crucial question, however, is from what dimensionality onward the improvement in fit only reflects error fitting. In terms of Kruskal's (1964) qualitative descriptions, for 8 of the 12 studied concepts, the seven-dimensional solutions can be situated between "fair" and "good." For the remaining 4 concepts, the stress value is in between the boundaries for "good" and "excellent" solutions. Furthermore, keeping in mind Kruskal's rule of thumb regarding the minimum number of stimuli needed for a solution in particular dimensionality, 5 (for *reptiles*, with 22 stimuli) up to 7 dimensions (for the 8 concepts with 30 stimuli) is the maximum number that we should be willing to accept. Second, taking the relative stress values as a choice criterion, determining the appropriate dimensionality is not easy. Clear elbows were seldom found and our best guesses led us to select 3 or 4 underlying dimensions for the vast majority of the studied concepts. (Note that 3 or 4 dimensions is even rather high, given that the literature search showed two-dimensional solutions in the vast majority of studies. Two possible reasons for this may be the relatively large number of stimuli we used, and the importance that was placed on interpretability and simplicity in the reviewed literature.) Third, evaluating the fit relative to the (estimated) reliability of the input similarity data resulted in a wider variety of underlying dimensionalities, with the number of retained dimensions varying from 2 to 6. And fourth, the chosen dimensionalities based on the Monte Carlo study also varied considerably (between 2 and 5), but the choices were not very similar to those based on the estimated reliability of the ratings.

In conclusion, the different choice procedures in which the dimensionality choice is (directly or indirectly) based on fit measures clearly result in inconsistent dimensionality choices.

The analyses in which the typicalities were correlated with distances toward the category centroid also yielded solutions in very diverse underlying dimensionalities in the different studied concepts (varying from 2 to 7). Note also that in many of the studied concepts, 6 or 7 underlying dimensions resulted in optimal predictions. This finding is in sharp contrast with the dimensionality of solutions

presented in the literature, where a dimensionality larger than 3 seldom occurs. In 9 of the 12 studied concepts, the predictive correlation pointed at a dimensionality that was higher than or equal to the selected dimensionality based on the other (traditional) choice criteria. We remind the reader that a dimensionality choice based on the correlation of an external criterion (i.e., a criterion other than the input similarities) cannot be inflated by error fitting. If retaining more dimensions results in an increase of the correlation, this proves that these additional dimensions contain semantic information that was not present in the solution offered in a lower dimensionality. And though the correlation improvement may be small with additional dimensions, the improved representation may reflect nuances that participants may find important. One might even ask oneself whether a representation consisting of more than seven dimensions yields even higher correlations. However, given the limited number of stimuli (22 to 30) that were used in Study 1, eight-dimensional solutions may not be warranted.

## STUDY 2

The analyses presented in the first study suggested that the traditional dimensionality choice criteria mostly yield inconsistent dimensionality choices. More importantly, a comparison of the results of these methods with those based on the prediction of an external criterion also showed that the traditional criteria sometimes underestimate the richness of the data.

In Study 2, we investigated whether—opposite the practice of MDS use in psychological applications—high dimensionalities can be informative about the semantics of the stimulus set. For this purpose, we used larger stimulus sets and categories that are less clearly delineated than those used in Study 1. More specifically, the stimulus sets consisted of larger semantic domains that cover several, possibly overlapping semantic concepts. Due to the vagueness of the boundaries of these concepts, judging typicalities for every stimulus within each of the different concepts is a sensible task and the centroid typicality prediction method could be used to investigate whether representations in even more dimensions than seven were appropriate.

The categories in Study 2 originated from the domain of artifacts and were derived from a previous naming study with common household objects (Ameel, Storms, Malt, & Sloman, 2005). In that study, participants named two sets of objects, which we will call bottles and dishes. The bottles set contained 73 pictures of objects that were selected to be likely to receive the name *bottle* or *jar* in American English, or else to share one or more salient properties with bottles and jars. For the dishes set, 67 pictures were selected of objects that were likely to be called *dish*, *plate*, or *bowl* in American English, or else, to have one or more salient properties in common with dishes, plates, or bowls.

Unlike in Study 1, typicality was predicted for multiple names applicable to large sets of stimuli in the studied domain. These names were selected such that they were

dominant for at least ten objects of a stimulus set. The three dominant names retained for the bottles set, were *fles*, *bus*, and *pot*. For the dishes set, *kom*, *tas*, *schaal*, and *bord* were selected. The categories from the bottles set contained, respectively, 25, 16, and 13 objects, the categories from the dishes set consisted of, respectively, 19, 15, 13, and 8 objects.

The typicality ratings of the two sets were taken from Ameel and Storms (2006), the similarity ratings from Ameel et al. (2005). For each object set, the exemplars were rated for typicality within the different categories of the set on a rating scale, ranging from 1 for *very atypical or unrelated items* to 7 for *very typical items*. Concretely, participants were asked to rate how typical each of the 73 (randomly presented) objects of the bottles set was for the categories *fles*, *bus*, and *pot*. Next, they judged the typicality of the 67 objects of the dishes set for the categories *kom*, *tas*, *schaal*, and *bord*. (Even though some of the stimuli were never named with a particular label, judging the typicality of these stimuli for these categories is perceived as sensible by the participants because the boundaries between the categories corresponding to the different labels are not clear-cut.) The order of presentation of the categories to be rated for an object set, as well as the order of presentation of the object sets, was counterbalanced across participants.

Because of the large number of objects in each of the object sets, Ameel et al. (2005) did not derive similarity matrices from directly rated pairwise similarities among all possible exemplar pairs. Instead they asked participants to sort the objects into piles based on overall similarity. The participants were instructed to provide as many piles as they wanted, but less than the number of stimuli and more than one single pile. Pairwise similarity was then recovered by counting for each of the object pairs the number of participants who placed that pair in the same pile. A large number of participants placing the two objects in a pile indicates high perceived similarity between the objects of the pair.

Reliabilities were estimated by the split-half correlations corrected with the Spearman-Brown formula. For the typicality ratings, the reliability estimates varied from 0.93 to 0.98 for the bottles and from 0.97 to 0.99 for the dishes. For the similarities, the estimates were, respectively, 0.93 and 0.92 for the bottles and dishes set.

## MDS Analysis and Dimensionality Tests

The similarity matrices for the two semantic domains were again analyzed with the MDS procedure of SAS (SAS, V8), using a nonmetric procedure. Following the rule of thumb that the number of stimuli should be larger than four times the number of dimensions, solutions in 2 to 18 and 2 to 16 dimensions were computed for, respectively, the bottles and the dishes set. Panels A and B in Figure 2 show, in solid lines, the stress values as a function of the underlying dimensionality, for the bottles and the dishes set, respectively.

If we would require a solution that would be qualified as “fair” by the Kruskal (1964) guidelines (i.e., stress value of .10), dimensionalities of 5 and 4 would be chosen for

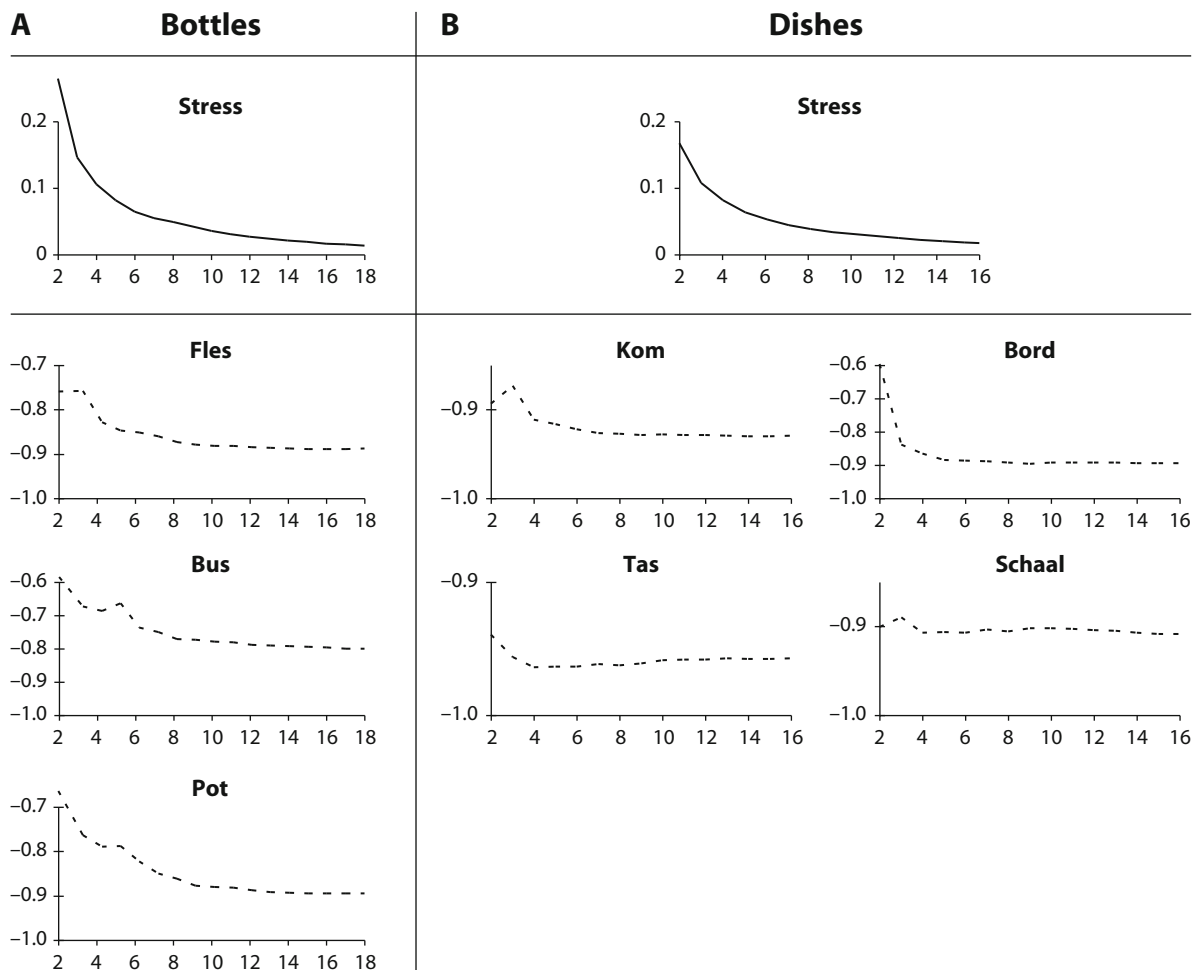


Figure 2. Stress values (solid lines) for MDS solutions in 2 to 18 dimensions for the bottles set (A) and in 2 to 16 dimensions for the dishes set (B) and predictive correlations (dashed lines) for the different concepts of the bottles set (A) and the dishes set (B) as a function of the underlying dimensionality.

the bottles and the dishes set, respectively. Requiring the solutions to be “good” according to the same guidelines (i.e., a stress value of .05), the chosen dimensionalities would be 8 and 6, respectively.

When evaluating the quality of the MDS solutions based on the relative stress values, the elbow pointed at the configuration in three dimensions for both the bottles and dishes set. Further, we determined the dimensionality of the solutions whose stress values best matched 1.0 minus the reliability of the input similarities. For the bottles set, the configuration in six dimensions was chosen and for the dishes set, the configuration in four dimensions.<sup>2</sup>

### Centroid Predictions of Typicality

As in Study 1, centroids for each category within a stimulus set were calculated by averaging, for every dimension, the coordinates of the exemplars of the category. Further, the Euclidean distances of each of the exemplars to the category centroids were calculated and correlated with the corresponding rated typicalities taken from

Ameel and Storms (2006). The predictive correlations for the categories within each stimulus set are shown, in dashed lines, in panels A and B of Figure 2, for the bottles and the dishes sets, respectively.

The correlations systematically increased with the underlying dimensionality. For the bottles set (Figure 2A), we found an increase up to dimension 15, 17, and 17 for, respectively, the categories *fles*, *bus*, and *pot*. For the dishes set, the correlations also increased up to very high dimensions for the categories *kom*, *schaal*, and *bord* (respectively 14, 16, and 14). Only for *tas*, distances computed from a four-dimensional solution yielded a correlation of  $-0.96$ , which did not further increase when adding a fifth dimension.<sup>3</sup> However, this is not that surprising since such a high correlation may indicate a ceiling effect.

Following the procedure of Study 1, we also computed distances to the central tendency points defined by the median value on each of the underlying dimensions. Next, the correlations between these distances and the typicalities in the category were calculated. Only for *fles* and *bord*,



the median-based predictions yielded a (nonsignificantly) higher correlation than those based on the centroid.

### Discussion

In Study 2, large stimulus sets were used. The results of the traditional dimensionality choice criteria again yielded relatively low dimensionality choices. On the contrary, when using an external criterion, for larger stimulus sets such as the ones used in Study 2, even more than seven dimensions were needed to optimally predict typicality. In fact, the correlations increased until dimensionalities 14 to 18. To our knowledge, solutions in so many dimensions have never been described in applications of MDS in the psychological literature. Although the improvements in the correlations are rather small in high dimensionalities, the increases cannot be an artifact of error fitting and they show that in order to explain typicality ratings in semantic concepts a very large number of dimensions is needed to describe all the subtle details that are present in the input similarities. Especially for the bottles set, correlations can be improved by .04 to .05 when allowing more than seven dimensions.

## GENERAL DISCUSSION

We set out to compare different dimensionality selection procedures for multidimensional scaling solutions of semantic concepts. Four traditional dimensionality choice procedures based on some measure of model fit and a new one in which the prediction of an external criterion was used were applied to two different data sets. For both sets it was found that the fit-based methods that are traditionally used to select the appropriate dimensionality of an MDS solution (absolute stress evaluation, relative stress evaluation, expected fit based on the reliability of the data, and a Monte Carlo-based procedure) often result in inconsistent choices. This finding casts doubt on the solutions discussed in the cognitive literature (described earlier in this paper), where these traditional choice procedures were used. Moreover, it was found that where these traditional methods of dimensionality selection prefer low-dimensional solutions, predictions of an external criterion sometimes show systematic information value in much higher dimensionalities. This detailed but meaningful information in similarity ratings among exemplars of semantic concepts is shown by the small but consistent predictive improvements in higher dimensionalities.

Our results suggest that commonly used fit-based measures of dimensionality selection often underestimate the number of sensible underlying dimensions. When one applies the prediction of an external variable like rated typicality as a criterion, systematic information can become apparent in very high dimensionalities. Besides these two classes of ways to determine the dimensionality of an MDS solution, we introduced a third one in the beginning of the paper. A priori theoretical reasons and interpretability proved to be rather important as indicated by their frequent application in the study of semantic concepts: Our review of the literature showed that in 37 of 138 analyses, one of these criteria was said to have de-

termined dimensionality selection. With an increase in dimensionality, as suggested by our results, the interpretation of the withheld dimensions becomes more difficult however. Many a researcher will probably think this to be an important drawback, considering the ease with which (traditional) MDS solutions could be interpreted as opposed to solutions of data analysis techniques as factor analysis.<sup>4</sup> However, when practical applications demand the identification of the dimensions that constitute a solution, regression techniques might be of help, providing a theory about the modeled content exists (Kruskal & Wish, 1978). Especially when one looks for more in MDS than a mere means of reducing and inspecting rich data sets, and one regards it as an appropriate way to undertake cognitive modeling (Lee, 2001), it might be important to rely on both data (fit and prediction of external variables) and theory (including interpretability).

The problem of dimensionality selection and interpretability of high-dimensional representations is not limited to applications of MDS. It is a problem faced by many models of semantic knowledge and often is a source of much debate. Latent semantic analysis (LSA, Landauer & Dumais, 1997) and the hyperspace analogue to language (HAL, Lund, Burgess, & Atchley, 1995; Lund & Burgess, 1996) are related to the MDS application discussed here in that they provide a high-dimensional spatial representation of semantic knowledge and hence face similar problems of dimensionality selection and interpretability. Typically, in applications of these models the very rich information in word document co-occurrence matrices is reduced to representation vectors of a dimensionality appropriate to capture the semantic information in the text files to which the models are applied. Although the algorithms for dimensionality *reduction* might differ from the MDS ones, related procedures of dimensionality *selection* may be applied. Landauer and Dumais (1997), for instance, after applying singular value decomposition, decided upon 300 as the appropriate number of dimensions to capture the similarity relations among the words of approximately 30,000 text samples. This choice of dimensionality was based on LSA's proportion of correct answers on an 80 item multiple-choice synonym test. When considerably more or less dimensions were withheld the model's fit to this external criterion was considerably impoverished. Of course, because of the high dimensionality of the retained solution one will not be surprised that LSA faces the issue of a problematic interpretability. A thorough account of ways to resolve this issue is not conceivable within the scope of this paper, but we refer the interested reader to work by Griffiths and Steyvers (2002) and Steyvers and Griffiths (in press) on probabilistic topic models in which this issue is explicitly addressed using an approach that is formally related to the LSA one.

The above described conclusions are mainly based on the (high) correlations obtained when predicting typicalities based on the distance toward centroids of the represented stimuli in varying dimensionalities. We note, however, that the comparisons reported may even have yielded underestimations of the number of dimensions needed to represent the systematic semantic information contained

in the input data. For one thing, in about half of the concepts investigated in Study 1, the correlations increased till the maximum dimensionality that can be allowed given the number of stimuli in these analyses. Given larger stimulus sets for these semantic concepts, correlations based on higher dimensionalities may still increase.

Furthermore, the analyses of Study 2 show that very high-dimensional solutions yield better predictive correlations than solutions in a lower dimensionality, but these results only show the value of additional dimensions *to the extent that the centroid allows one to predict typicality*. In other words, the predictive correlations depend heavily on a number of assumptions, which are theoretically defensible, but which may not all be completely true. For instance, it is not sure whether the optimal prediction point for typicality is the centroid of the exemplar cloud. The results of Study 1 suggest that median-based central tendencies may be better suited as typicality predictors in some concepts. (For evidence in favor of a caricature ideal point that is to some extent influenced by possible contrast categories, see Ameel & Storms, submitted.)

The large number of dimensions needed to extract the systematic information included in the input similarities also casts doubt on the optimality of MDS as a data-analytic technique to analyze input similarities in the domain of semantic concepts. It is conceivable that many of these dimensions have a limited number of dimension values, and that they can better be represented with dichotomous features that are either present or absent in a particular stimulus. Suppose, for instance, that the feature “has wings” co-determines the input similarities to a large extent. MDS is able to represent this by adding a wing dimension to the solution. However, the stimuli will only have two different coordinate values on this underlying dimension. If many of the underlying features have a dichotomous—rather than a continuous—nature then feature models, like additive trees (De Soete, 1983) or additive clustering (Shepard & Arabie, 1979) may provide more parsimonious and even better fitting accounts of the similarity data. (For evidence in line with this hypothesis, see Pruzansky, Tversky, & Carroll, 1982, and Tversky & Hutchinson, 1986.) Further study, in which these models are applied to semantic data, is needed to investigate this question. In such studies, both model fit and prediction of external variables (like, e.g., typicality) may be investigated and compared.

Finally, after having selected the dimensionality with the external typicality prediction criterion, the chosen representations could be used to account for other external semantic data, like, e.g., categorization, reaction times, induction, etc.<sup>5</sup> More research is needed to further explore the relation between the dimensionality of MDS solutions and such other external variables.

#### AUTHOR NOTE

The research presented in this paper was partly sponsored by grants OT/05/27 and IDO/02/004 of the Leuven University Research Council awarded to Gert Storms. The two first authors contributed equally to this manuscript. Eef Ameel is research assistant at the Fund for Scientific Research, Flanders. Authors' addresses: Steven Verheyen and Eef

Ameel, Department of Psychology, Tiensestraat 102, B3000 Leuven, Belgium (e-mail: steven.verheyen@psy.kuleuven.be and eef.ameel@psy.kuleuven.be).

#### REFERENCES

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE: Transactions & Automatic Control*, **19**, 716-723.
- AMEEL, E., & STORMS, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory & Language*, **55**, 402-421.
- AMEEL, E., STORMS, G., MALT, B. C., & SLOMAN, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory & Language*, **53**, 60-80.
- ARNOLD, J. B. (1971). A multidimensional scaling study of semantic distance. *Journal of Experimental Psychology*, **90**, 349-372.
- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.
- BARSALOU, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition: Content and process specificity in the effects of prior experiences* (Vol. 3, pp. 61-88). Hillsdale, NJ: Erlbaum.
- BORG, I., & GROENEN, P. J. F. (1997). *Modern multidimensional scaling*. New York: Springer.
- DE SOETE, G. (1983). A least squares algorithm for fitting additive trees to proximity data. *Psychometrika*, **48**, 621-626.
- GÄRDENFORS, P. (2004). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.
- GRIFFITHS, T. L., & STEYVERS, M. (2002). A probabilistic approach to semantic representation. In W. D. Gray, & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 381-386). Mahwah, NJ: Erlbaum.
- 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 representations. 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.
- HOLYOAK, K. J., & WALKER, J. H. (1976). Subjective magnitude information in semantic orderings. *Journal of Verbal Learning & Verbal Behavior*, **15**, 287-299.
- JONES, R. A., & ROSENBERG, S. (1974). Structural representations of naturalistic descriptions of personality. *Multivariate Behavioral Research*, **9**, 218-230.
- KOMATSU, L. K. (1992). Recent Views of conceptual structure. *Psychological Bulletin*, **112**, 500-526.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.
- KRUSKAL, J. B., & WISH, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.
- LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.
- LEE, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, **45**, 149-166.
- LUND, K., & BURGESS, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, **28**, 203-208.
- LUND, K., & BURGESS, C., & ATCHLEY, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In J. D. Moore, & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 660-665). Hillsdale, NJ: Erlbaum.
- McKINLEY, S. C., & NOSOFKY, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.
- MEDIN, D. M., & 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*, *5*, 355-368.
- NOSOFSKY, R. M. (1984). Choice, similarity, and the context model 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.
- PRUZANSKY, S., TVERSKY, A., & CARROLL, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika*, *47*, 3-24.
- RAMSAY, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, *42*, 241-266.
- RAMSAY, J. O. (1980). Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika*, *45*, 139-144.
- 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.
- RIPS, L. J., SHOBEN, E. J., & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, *12*, 1-20.
- ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.
- ROSENBERG, S., & JONES, R. (1972). A method for investigating and representing a person's implicit theory of personality: Theodore Dreiser's view of people. *Journal of Personality & Social Psychology*, *22*, 372-386.
- RUTS, W., DE DEYNE, S., AMEEL, E., VAN PAEMEL, W., VERBEEMEN, T., & STORMS, G. (2004). Flemish norm data for 13 natural concepts and 343 exemplars. *Behavior Research Methods, Instruments, & Computers*, *36*, 506-515.
- RUTS, W., STORMS, G., & HAMPTON, J. (2004). Linear separability in superordinate natural language concepts. *Memory & Cognition*, *32*, 83-95.
- SAS INSTITUTE INC. (1999). *SAS STAT Users' Guide 8*. Cary, NC: Author.
- SCHIFFMAN, S. S., REYNOLDS, M. L., & YOUNG, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications*. New York: Academic Press.
- SCHWARZ, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, *6*, 461-464.
- SHEPARD, R. N., & ARABIE, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, *86*, 87-123.
- SHOBEN, E. J. (1976). The verification of semantic relations in a same-different paradigm: An asymmetry in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, *15*, 365-379.
- SHOBEN, E. J. (1983). Applications of multidimensional scaling in cognitive psychology. *Applied Psychological Measurement*, *7*, 473-490.
- SHOBEN, E. J., & ROSS, B. H. (1987). Structure and process in cognitive psychology using multidimensional scaling and related techniques. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing*. London: Erlbaum.
- 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.
- SMITH, J. D., & MINDA, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, & 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 & Review*, *9*, 836-844.
- SPENCE, I. (1983). Monte Carlo simulation studies. *Applied Psychological Measurement*, *7*, 405-425.
- SPENCE, I., & GRAEF, J. (1974). The determination of the underlying dimensionality of an empirically obtained matrix of proximities. *Multivariate Behavioral Research*, *9*, 331-341.
- STEYVERS, M., & GRIFFITHS, T. L. (in press). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A road to meaning*. Mahwah, NJ: Erlbaum.
- STORMS, G. (1995). On the robustness of maximum likelihood scaling for violations of the error model. *Psychometrika*, *60*, 247-258.
- STORMS, G., DIRIKX, T., SAERENS, J., VERSTRAETEN, S., & DE DEYN, P. P. (2003). On the use of scaling and clustering in the study of semantic disruptions. *Neuropsychology*, *17*, 289-301.
- TAKANE, Y. (1981). Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, *46*, 9-28.
- TAKANE, Y., & CARROLL, J. D. (1982). Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, *46*, 389-405.
- TVERSKY, A., & HUTCHINSON, W. (1986). Nearest-neighbor analysis of psychological spaces. *Psychological Review*, *93*, 3-22.
- WAGENAAR, W. A., & PADMOS, P. (1971). Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. *British Journal of Mathematical & Statistical Psychology*, *24*, 101-110.

## NOTES

1. We performed significance tests for each pair of successive correlations. However, we did not find any systematic results. For *birds* and *insects*, the increase in correlations was significant up to dimension 6, for the other concepts, only a few pairs of successive correlations were significantly different, regardless of whether the highest correlation was included in the pair or not. For 7 out of the 12 concepts, increases corresponding to comparisons of lower dimensionalities were not significant, while increases corresponding to comparisons of higher dimensionalities turned out to be significant. These results suggest that looking at statistical significance in the pattern of correlations is not a very useful strategy to determine whether additional dimensions account for unique variance.
2. A Monte Carlo procedure, like the one in Study 1, in which the choice procedure described by Wagenaar and Padmos (1971) is employed, was not practically possible for the data set used in Study 2. Factorially combining true underlying dimensionalities varying between 2 and 18 with solutions obtained in dimensionalities 2 to 18, with 7 different error levels, and with 100 replications for each combination, would, for both studied concepts, lead to more than 400,000 analyses.
3. As in Study 1, significance tests were performed. Although more systematic results were found for Study 2 (e.g., the highest correlation was always significantly higher than the correlation before it, except for *kom* and *schaal*), the results again showed an irregular pattern (e.g., significant increases still alternated with nonsignificant increases and increases in high dimensions were more often significant than increases in low dimensions).
4. Schiffman, Reynolds, and Young (1981) note that this advantage is partly due to the fact that it comes more natural to people to interpret distances between points than angles between vectors, on which the solutions in factor analysis are based. They also note that the factor analysis solutions generally consist of more dimensions than the MDS ones, making them harder to interpret.
5. We thank an anonymous reviewer for this suggestion.

**APPENDIX**

The Monte Carlo simulation studies were performed as follows. First, spatial configurations of stimuli in a particular dimensionality were constructed by randomly sampling the coordinates of the stimuli on all dimensions from a uniform distribution between 0 and 1000. After computing the Euclidean distances between the stimuli, these distances were error perturbed by multiplication of each errorless distance with a random factor, drawn from a normal distribution with a mean of 1.0 and a particular variance (0, 0.05, 0.10, 0.15, 0.20, 0.25, or 0.30). This procedure was repeated 100 times for each of the concepts, with a number of (randomly localized) stimuli that was equal to the number of stimuli in the data set under investigation. Besides the error level of the generated data (i.e., the variance of the error component), the underlying true dimensionality of the simulated data was also varied: Stimuli were generated in 2 to 7 dimensions.

The randomly generated dissimilarity sets were then used as input for the SAS MDS procedure. For every combination of the true underlying dimensionality, the dimensionality of the solution, and the error level, the stress values of the solutions for all 100 replications were (1) averaged and (2) rank ordered. Next, the appropriate dimensionality was determined using a procedure similar (but not identical) to the Wagenaar and Padmos (1971) method. More specifically, the procedure always started with the assumption that a representation in two dimensions (2D) was the most appropriate. Based on this assumption, the error level of the data was determined: The stress value of the data in 2D was compared with the average stress value (over the 100 replications) of the 2D solutions for truly 2D simulated data. If the stress value of the empirical data was in between the averaged stress for two different error levels (e.g., error variances .10 and .15), then an exact estimate of the error level was determined through linear interpolation between these two bordering error levels. Then a nonmetric decision procedure was applied to test the hypothesis of two underlying dimensions. First, the percentile 5 values of the stress values corresponding to 3D solutions for the error distributions of the two bordering error levels were selected. Second, a value in between these two figures (i.e., in between the fifth percentile values of the distributions of the two bordering error levels) was calculated using the same interpolation position that was obtained in the exact error level determination. Finally, if the stress value of the 3D solution for the empirical data was larger than the expected (interpolated) fifth percentile stress value, then the hypothesis of two underlying dimensions was retained; if the value was lower than the expected fifth percentile, then the hypothesis of a 2D underlying solution was rejected. In the latter case, the whole procedure was started over again, assuming an underlying 3D representation. The procedure was repeated until the stress of the empirical data exceeded the fifth percentile of the stress value distribution corresponding to the dimensionality just above the hypothesized dimensionality

---

(Manuscript received December 3, 2005;  
revision accepted for publication April 26, 2006.)