Psychologica Belgica
2012, 52/4, 435-458

AN EXEMPLAR APPROACH TO CONCEPTUAL COMBINATION

Wouter Voorspoels, Gert Storms, &
Wolf Vanpaemel*
KU Leuven

Despite the success of exemplar models of representation, the general approach in the study of the representation of conceptual combinations is based on prototypes. In this study, we evaluate the exemplar view in conceptual combination, and compare it to the traditional prototype approach. For 10 complex concepts, typicality was predicted using an instantiation-based spatial exemplar model. The exemplar model's predictions were compared to the predictions of five plausible prototype models. Results clearly indicated that the exemplar model provides the best predictions of typicality for the complex concepts, and little or no unique variance in the observed typicality gradient was accounted for exclusively by the prototype model. We find that an exemplar representation of five to ten exemplars is optimal in the prediction of typicality, which is remarkably similar to earlier findings regarding simple, established concepts. Following the clear dominance of exemplar models in categorisation and concept research, the present study underlines the need of taking the exemplar approach to the next level, applying it in the more complex field of conceptual combination.

Introduction

An intriguing aspect of human language is the flexibility people display in the use and interpretation of natural language concepts in everyday language use. People are seemingly able to effortlessly adapt, combine or specify semantic concepts – such as 'sports', 'weapons' or 'vehicles' – to communicate more accurately what is intended (e.g., 'indoor sports', 'weapons used in war' or 'vehicles for transporting people'). Language comprehension and production seem to necessarily imply "… the combination of concepts into larger and larger structures as guided by the syntax of language" (Murphy, 2002, p. 443). As an illustration of this point, we present the fol-

* Wouter Voorspoels, PhD, Gert Storms, PhD, Wolf Vanpaemel, PhD, Department of Psychology, KU Leuven.
The research reported in this article is part of an interdisciplinary research project (IDO/02/004) and research project OT/05/27 sponsored by the Research Council of the university of Leuven. Both of these grants were given to the second author.
Correspondence concerning this article should be addressed to Wouter Voorspoels, KU Leuven Department of Psychology, Concat, Tiensestraat 102, B-3000 Leuven. Email: wouter.voorspoels@psy.kuleuven.be

lowing randomly[1] chosen extract from a newspaper:

> "…
> The agreement will also strengthen Google's dominance over the ***lucrative search advertising market***. It was signed after Yahoo rejected a proposal by Microsoft to acquire both Yahoo's ***search business*** and a ***minority stake*** in the company. The rejection appears to end months of ***on-again, off-again negotiations*** between the two companies.
> …"
> (Helft, 2008, *New York Times*, June 13, 2008, bold italics added).

Just looking at rather small "larger and larger structures" of only a few words, we can count 4 concept combinations in this short *New York Times* paragraph. And leaving the present sentence out of the count, in what we have written up until now we used 15 different combinations. Each combination can be seen as a specific semantic category, which is constructed from constituent concepts, to fit as best as possible the intended meaning (Wisniewski, 1998).

Compositionality and productivity, that is to say, the ability to form new concepts on the basis of acquired concepts should be a core topic in concept research (e.g., Fodor & Lepore, 2002), and moreover, it is an important test of the generality of theories on natural language concepts. How do people arrive at constructing and interpreting complex concepts, such as 'homicidal green penguin' (Osherson & Smith, 1981)? While context and language syntax undoubtedly have some role in interpreting these larger structures, it is obvious that the interpretation of the combination of relatively simple concepts into more complex concepts is for a large part determined by the meaning, and thus the representation, of the relatively simple[2] concepts 'penguin', 'homicidal' and 'green' (e.g., Hampton, 1997). In the present study, we examine this issue in more detail.

There exist different lines of research that tap into the domain of conceptual combination. Broadly speaking, two perspectives can be identified. A first perspective focuses on the different strategies that people apply when interpreting combinations of concepts. For example, a number of researchers have examined the *critical relations* that allow the interpretation of noun-noun combinations (e.g., Gagné & Shoben, 1997; Levi, 1978; Wisniewski & Murphy, 2005), such as 'mountain stream' (a stream LOCATED AT a mountain) and 'mountain magazine' (a magazine ABOUT mountains). The selection of the appropriate relation (that is, the relation that is generally chosen by

1. In fact, the paragraph was not chosen randomly, but was selected to illustrate the point. In all honesty, we can say that the search did not take a long time.
2. We use the term 'simple' to denote concepts for which a well established, lexicalised expression exists. Complex concepts are concepts build from several simple concepts.

language users) has been found to be dependent on the distributional information that people have regarding the frequency of association of a particular relation with a concept: For example, the concept mountain often elicits the selection of the LOCATED AT relation (Gagné & Shoben, 1997; but see, Devreux & Costello, 2007; Maguire, Devereux, Costello, & Cater, 2007).

For some combinations the interpretation does not follow from the selection of a critical relation. Rather, the selection of an appropriate property is crucial (Wisniewski, 1998). For example, in the combination 'black chair' – the property 'black' is quite straightforwardly mapped on the concept 'chair'. Sometimes, for example in the combination 'zebra chair', the choice of property is less straightforward. In this case, participants generally interpret the combination by mapping the property 'striped' onto 'chair' (e.g., Estes & Glucksberg, 2000; Wisniewski & Middleton, 2002). The question then is what determines the preference to map a particular property over another.

A second research perspective, which is taken in the present study, has focused on how language users determine the membership of a newly formed complex concept, and more particularly, the manner in which the membership structure of the complex concept can be deduced from the membership structures of the simple concepts involved (Costello, 2000; Hampton, 1988; Murphy, 1990; Smith & Osherson, 1984). For example, a shark is an instance of the concept 'fish', and of the concept 'predator'. As such it is a likely candidate to be a member of the complex concept 'predator fish'. This perspective nicely aligns with a large body of research regarding categorisation and concepts. The main aim is to generalise the theories that stem from the research on simple concepts, to the domain of complex concepts. In what follows, we discuss the dominant theoretical framework taken within this perspective and point out two challenges that suggest the need for a shift in framework.

Challenges to a prototype view of complex concepts

Following the main approach in research concerning natural language categories, theories of conceptual combination are traditionally based on a prototype view on concepts. In this view it is assumed that simple, semantic concepts are represented by a prototype, that is, a summary representation often assumed to be the average of the category (e.g., Hampton, 1993; Minda & Smith, 2010; Posner & Keele, 1968; Younger & Cohen, 1983). The concept 'weapons', for example, is assumed to be a summary representation of what weapons are like on average.

Extending this approach to the domain of conceptual combination, several models have been developed that use this notion of a prototype to give an account of how people interpret complex concepts such as 'dangerous weap-

ons' or 'red apple' (Murphy, 1990; Smith, Osherson, Rips, & Keane, 1988). In these models, a concept is typically seen as a schema, consisting of dimensions (e.g., colour, shape, size) and possible values on these dimensions. The schema representation of a concept such as apple may contain the dimensions colour, shape, texture and size. The dimension for colour would contain possible values such as 'red', 'green' and 'brown', each of which has a certain salience within the concept. When the concept 'apple' is combined with another concept to form for example the complex concept 'red apple', the dimension 'colour' becomes dominated by the value 'red', and the dimension of colour is weighted more heavily. The net result is that the dimension 'colour' becomes more diagnostic in determining whether something is a red apple than the dimension colour would be in a judgment of whether something is an apple. In short, the conceptual combination 'red apple' results in a modification – essentially a reweighting of features – of the prototype of the concept of 'apple'.

There are however two major challenges for these prototype models of complex concepts. First, several intuitions and observations suggest that the extension of complex concepts – i.e., the set of things in the world the concept refers to – influences the representation (e.g., Gray & Smith, 1995; Hampton, 1997; Medin & Shoben, 1988; Murphy, 1990). For example, Medin and Shoben (1988) have shown that a metal spoon is judged to be more typical of 'spoons' than a wooden spoon, whereas a wooden spoon is judged more typical of 'large spoons' than a metal spoon. This is problematic for the prototype based models since there is no *a priori* reason why modifying the size dimension of the concept 'spoon' affects the salience of a certain value on another dimension. However, many instances of the category 'large spoon' are made of wood, and it seems people use their knowledge of stored instances of the concept 'large spoon' to judge typicality (see also Gray & Smith, 1995). The influence of extensional information when we combine concepts is often referred to as *extensional feedback* (e.g., Hampton, 1997). Despite clear evidence for such influence, well specified and empirically grounded ways of implementing extensional feedback in models of conceptual combination are rare (for exceptions, see Costello, 2001; Devereux & Costello, 2007).

A second and perhaps even greater challenge for prototype models of conceptual combination, is the rise of exemplar models of representation. According to the exemplar view, categories are represented by previously encountered instances of a category, rather than an abstracted prototype (e.g., Brooks, 1978; Medin & Schaffer, 1978). The concept 'weapons' thus is assumed to be represented by members of the category. The exemplar approach has proven successful in a large array of conditions, encompassing both artificial category learning (e.g., Busemeyer, Dewey, & Medin, 1984; Nosofsky, 1992; Vanpaemel & Storms, 2010) and natural language concepts

(Voorspoels, Vanpaemel, & Storms, 2008), and in general compares favourably to the prototype approach. Obviously, these findings are problematic for the traditional models of conceptual combinations, that have their roots in prototype representations.

Outline

Both the notion of extensional feedback and the success of the exemplar view in studies concerning simple concepts, point to the necessity of a thorough evaluation of the role of exemplar information in conceptual combination, starting from recent models used in simple concept research. In this study, we implement and evaluate exemplar representations as representation of complex concepts, and contrast the model with a number of traditional prototype models. We compare the models in their account of the typicality gradient of complex concepts.

The notion of typicality refers to the observation that some members of a category are better examples of the category than are others. Cows are generally seen as more typical examples of the category 'mammals' than are duckbilled platypuses, or whales. Typicality has been shown to be an influential variable in a wide range of cognitive tasks (for a review see Hampton, 1993), and one of the most important variables in semantic concept research. As such, typicality can be considered an important criterion in evaluating theories of concepts: a theory of concept representation that can not account for the typicality gradient is no good.

To evaluate and test the exemplar approach in the context of complex concepts, we derive typicality predictions for the members of complex concepts from a straightforward exemplar model. These predictions will then be compared to the predictions of a number of plausible prototype models in how well they approximate the observed typicality ratings. The key idea behind the exemplar model is that a complex concept is represented by a set of exemplars that is activated. This set of exemplars constitutes a subset of the members of the unmodified concept. The model predicts that highly typical members of a modified concept are similar to the set of exemplars that is activated. This model will be referred to as *spatial exemplar model*, due to its reliance on a spatial stimulus representation.

To challenge the exemplar account of the typicality gradient in complex concepts, we consider two types of prototype models. The first type is based on the idea of conceptual combination as the modification of a prototype in the sense of a reweighting of the features of the unmodified concept. A prototype model based on reweighting the features will be referred to as the *feature-based prototype model*. Intuitively, a feature-based prototype model predicts that highly typical members of a modified concept have the features that

are important for the modified concept (that is, features that have a high feature weight after reweighting). Different ways to arrive at appropriate weights for the features will be considered.

We also test a second type of prototype models, that is, based on a spatial stimulus representation as is the spatial exemplar model. In a *spatial prototype model*, a complex concept is represented by the central tendency of a subset of exemplars of the unmodified concept. The more an exemplar is similar to the subset's central tendency, the more typical it is of the modified concept. The central tendency can be the average or the median, and both will be tested.

In the next sections we will first give an overview of the data we used in the present study followed by a detailed overview of the exemplar model, a basic spatial prototype model and a basic feature-based prototype model. After this, we will present and discuss the results of the model evaluations. In the model comparisons, we will consider different variants of the prototype models.

Data

The dependent variable in this study is a measure of typicality. To derive the feature-based prototype measure of typicality for the complex concepts, we used previously published feature applicability ratings (De Deyne, Verheyen, Ameel, Vanpaemel, Dry, Voorspoels et al., 2008) and we collected additional feature importance ratings to determine the reweighting of features when modifying a concept. To obtain a spatial representation on which the exemplar model and the spatial prototype model is based, we used previously published similarity ratings (De Deyne et al., 2008). Finally, we also gathered categorisation judgments. These judgments are used in the *spatial exemplar model* and the *spatial prototype model*.

Stimulus set

Complex concepts were created starting from five common, simple natural language categories ('sports', 'musical instruments', 'vehicles', 'clothing', and 'weapons') taken from a recent norm study (De Deyne et al., 2008). Each of these categories contains between 20 and 30 (verbal) instances.

For each of the 5 common concepts, we construed two complex concepts, resulting in 10 complex concepts, which were specifications of the basic categories: 'indoor sports' and 'outdoor sports', 'musical instruments used in rock music' and 'musical instruments used in classical music', 'vehicles used for the transport of people' and 'vehicles used for the transport of goods', 'summer clothes' and 'winter clothes', 'weapons used in wars' and 'weapons

used for sports'[3]. The complex concepts contained at least some of the members of the simple concepts from which they were derived. For example, the simple concept 'sports' entails members such as 'basketball', 'volleyball' and 'ballet' – which intuitively are 'indoor sports' – but also members such as 'rugby', 'skiing' and 'sailing' – which intuitively are 'outdoor sports'.

Typicality ratings

We used a goodness-of-example[4] measure to assess the typicality of an instance for a category. All instances of each simple concept were rated for goodness-of-example for each associated complex concept by 20 to 26 participants. Reliabilities, estimated using split half correlations and corrected with the Spearman-Brown formula ranged from .91 to .98. A typicality score for each instance towards the relevant complex concept was obtained by averaging the typicality ratings across participants. These averaged typicality scores are used for further analysis.

Feature applicability and feature importance ratings

For each of the simple concepts, De Deyne et al. (2008) report an exemplar by feature matrix, containing between 32 and 39 features generated for the concepts. The matrices contain judgments – elicited from four participants – of the applicability of each feature for each exemplar of a simple concept. In other words, these matrices contain information on whether a particular member of the concept has a feature that was generated for the concept (e.g., whether 'basketball' has the feature 'generates transpiration', which was generated for the concept sport). The reliability of the applicability judgments per concept was evaluated applying Spearman-Brown formula to split-half correlations, resulting in estimated reliabilities between .83 and .88 (De Deyne et al., 2008).

We collected additional data capturing the importance of these features for the complex concepts. For each of the complex concepts in this study, we asked 10 to 15 participants to rate the importance of the relevant features[5]. For example, participants were asked to judge the importance of the feature 'generates transpiration' for the complex concept 'outdoor sports'. Applying the Spearman-Brown formula to the split-half correlations, all reliabilities except one were estimated between .81 and .93. For 'weapons used for

3. These are (free) translations of the stimuli that were actually used.
4. Typicality ratings and goodness-of-example ratings are both measures of graded structure in concepts. They are often seen as synonymous.
5. I.e., the features that were generated for the simple concept.

sports', the reliability was .64, which is rather low. These feature importance ratings will allow to reweight the features in the feature-based prototype model.

Similarity ratings and underlying representations

For the five simple concepts pairwise similarity ratings were available from the norm studies (De Deyne et al., 2008). For each category, all pairwise similarities were judged by 14 to 25 participants. Reliability of the ratings was evaluated using split half correlations, corrected with the Spearman-Brown formula, and ranged between .89 and .96. The averaged similarity matrices were used as input for MDS-analyses to arrive at underlying spatial stimulus representations.

Categorisation decisions

Using a simple computerized categorisation task, 35 participants were presented with the instances of a simple concept and were asked to indicate to which of the appropriate two complex concepts the instance belonged. The task thus consisted of five blocks, one for each simple concept, and each block consisted of all instances of a simple concept (thus ranging from 20 to 30 instances). In each trial, a fixation cross was presented in the middle of the screen, followed by the stimulus. The stimulus remained on the screen until an answer was given, for a maximum of 10 seconds. The order of presentation of the instances was random, as well as the order of the five blocks. Categorisation proportions were derived for each of the instances of a simple concept with respect to the appropriate complex concepts.

On the basis of the categorisation proportions, we derived the category sizes of the ten complex concepts. Applying a criterion of 80% agreement between participants, the category sizes range from 4 to 17, with a median of 9. Our set of complex concepts thus is not biased to small or large categories, which is important in the light of the influential role attributed to category size in the exemplar versus prototype debate (see e.g., Minda & Smith, 2001).

Models

Feature-based prototype model

In the feature-based prototype model the representation of a concept is assumed to consist of a set of (weighted) features. As noted earlier, prototype modification as proposed by traditional theories of conceptual combination (e.g., Smith et al., 1988) essentially comes down to a reweighting of the fea-

ture structure. Typicality of an instance towards the modified concept then is the similarity of the instance towards the (re-)weighted feature representation. This can easily be calculated by summing the importance of a feature multiplied by the degree to which a certain instance has this feature.

Formally, for an instance i with F features, the typicality towards complex concept A is given by:

$$T_{(i,A)} = \sum_{j=1}^{F} (I_{jA} \times T_{ji}) \quad (1)$$

in which I_{jA} is the weight of feature j for complex concept A, and T_{ji} is the applicability of feature j to instance i.

To implement the model, we use information on the applicability of a feature to an exemplar, and the importance of the feature for the modified concept. If an exemplar has the features that receive a strong weight for the modified concept, it is predicted to be typical of the modified concept.

Spatial models

The predictions of typicality of the spatial models are based on underlying spatial stimulus representations of the simple concepts from which the complex concepts are derived. In such similarity spaces, the instances of a category are represented as points in an M-dimensional space and the distance between two instances in the space is inversely related to the similarity between the instances. Depending on the model – a prototype or an exemplar model – typicality is translated as the distance (i.e., the inverse of similarity) towards the average point of a category, (i.e., the prototype), or the summed distance of the instance towards all other instances. Spatial models have already been proven to be quite successful in the representation of basic semantic concepts and more specifically in accounts of typicality (e.g., Verheyen, Ameel, & Storms, 2007; Voorspoels et al., 2008).

The concept representation of a complex concept was built using an instantiation process, in which a certain subset of exemplars in the underlying spatial representation is used. We will in turn describe the spatial prototype model, the exemplar model and the instantiation principle that is applied in both models.

Exemplar model

According to the exemplar view a concept representation consists of all members of a category. Typicality of an instance to a category then is the summed similarity of the instance towards all members of the category. For stimulus i with M dimensions, the typicality to complex concept A is predicted to be:

$$T_{(i,A)} = \sum_{j=1}^{n} \left(\sum_{k=1}^{M} (x_{ik} - x_{jk})^2 \right)^{1/2} \quad (2)$$

where the instances j are members of the set (of size n) that make up the category representation and x_{ik} is the coordinate of instance i on dimension k.

Prototype model

The most dominant implementation of a prototype is the average instance of the category (see, e.g., Minda & Smith, 2010) . Typicality of an instance to a category according to the prototype view is the similarity of that instance to the prototype. Formally, the predicted typicality of instance i to complex concept A is given by:

$$T_{(i,A)} = \left(\sum_{k=1}^{m} (x_{ik} - p_{Ak})^2 \right)^{1/2} \quad (3)$$

where x_{ik} is the coordinate of instance i on dimension k, p_{Ak} is the coordinate of the prototype of category A on dimension k and M is the number of dimensions of the underlying representation. The prototype is found by averaging across the coordinates of these instances on each dimension:

$$P_{Ak} = \frac{1}{n} \sum_{i=1}^{n} x_{ik} \quad (4)$$

in which i is an element of the set of instances, with size n, that are included in the representation of category A. Note that the instances included in the calculation of the prototype will determine the location of the prototype.

The instantiation principle

In semantic concept research, an instantiation principle has been proposed (Heit & Barsalou, 1996) that essentially states that for category decisions – such as categorisation decisions, but also typicality judgments – one (optimal) category member is activated. This principle is generalised in De Wilde, Vanoverberghe, Storms, and De Boeck (2003), such that an optimal subset of members of the category is activated instead of only one.

In the present study, both the prototype and the exemplar models require a specification of the exact set of category members that are included in the representation (see equation (2) and (4)). A process inspired by the instantiation principle is easily implemented in formulas (2) and (4) by choosing the number of instances included and the specific instances that are instantiated.

Based on the categorisation proportions, we made a ranking of instances for each complex concept in terms of the proportion of people that judged them as belonging to the category. For each complex concept we then selected the *n* (ranging from 2 to 20) instances which were most agreed upon to belong to the category (i.e., with the highest categorisation proportion for the category). The resulting set of *n* "optimal" instances was then used in the exemplar (equation 2) and prototype model (equation 4).

Results

Stimulus representations

We obtained an underlying spatial representation for each of the five simple concepts, using the pairwise similarity ratings as input for a SAS MDS analysis (SAS, V9). We considered solutions in 2 to 8 dimensions for all concepts. The quality of a solution can be evaluated through the stress, a measure of badness-of-fit. Stress values decreased monotonically as a function of dimensionality, indicating the routine did not get trapped in a local minimum for any of the solutions. For three concepts, stress values dropped below .1 from Dimensionality 4 onwards, for the remaining two, stress values dropped below .1 from Dimensionality 5.

Model comparison

The performance of the different models was assessed by computing the correlation between the empirically observed and the predicted typicality. For the feature-based spatial model, the predictions are based on the feature applicability and importance scores. For the models based on an underlying similarity space, predictors of typicality are dependent on the particular set of instances that are activated. We calculated typicality predictions including 2 to 20 instances – and this was done for underlying spatial representations in dimensionalities 4 to 8. For each dimensionality the optimal number of instances (i.e., resulting in the concept representation that produces the best correlation with observed typicality) was chosen.

In Figure 1 the performance of the models is presented, as a function of dimensionality. Since the feature-based prototype model is not based on the underlying spatial representation, it yields only one prediction for each complex concept, presented by the horizontal dashed line[6].

6. In 'weapons used for sports', the feature-based prototype model yielded a correlation close to zero, and was not added in the graph. This might be due to the low reliability (.62) of the feature importance ratings, which are essential in the calculation of this measure, and might point to confusion of the participants regarding this concept.

The spatial exemplar model (solid line) provides a better prediction of typicality than the feature-based prototype model (dashed line) in all but one category ('vehicles for transporting people'). For 'summer clothes', the spatial exemplar model outperforms the feature-based prototype model from Dimensionality 5 onwards. Compared to the spatial prototype model, the exemplar model consistently performs better in all categories and dimensionalities. In Dimensionality 8, the exemplar model produces an average correlation of .75 with observed typicality (averaged across categories), as compared to .52 for the feature-based prototype model, and .71 for the spatial prototype model.

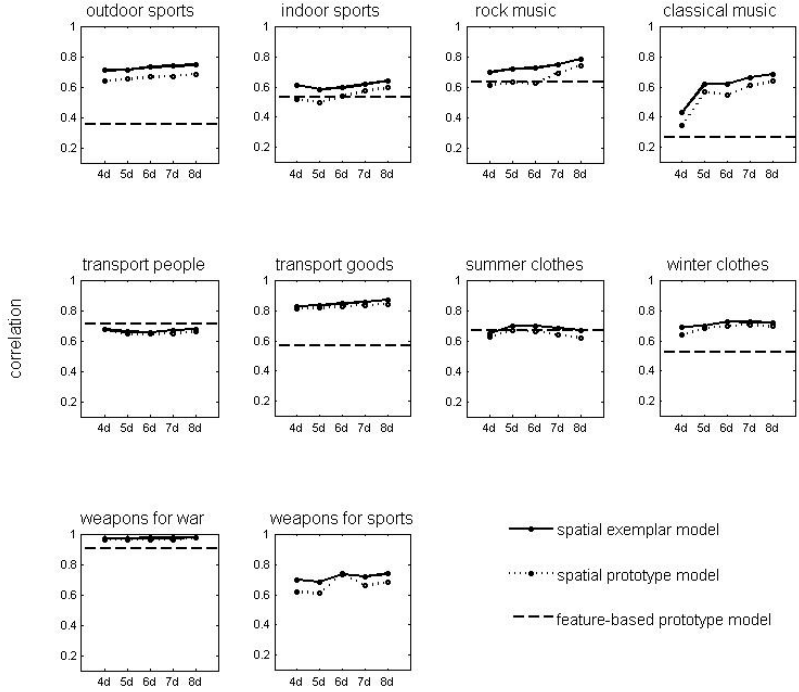


Figure 1
Correlation between observed and predicted typicality for the 3 models as a function of dimensionality

A potential concern in the comparison between the feature-based prototype model and the exemplar model is that the feature-based prototype model might have suffered from the lack of freedom available to the exemplar model. However, this difference is non-existent for the comparison between

the two spatial models. Note that the prototype model is based on the same underlying spatial representations, and has access to similar information to select a subset of instances. The only difference between these two models is that the exemplar model uses optimally selected instances as representation, and the prototype model averages over an optimally selected subset of instances. Figure 1 shows that the exemplar model also outperforms the spatial prototype model (dotted line) for the 10 complex concepts. While differences are rather small for some complex concepts, the exemplar model consistently predicts the observed typicality better.

Apart from looking at the performance of each model separately, it is also worthwhile to investigate whether the exemplar and the prototype models capture a different aspect of the variability in typicality ratings. It might be that some important aspect of the typicality gradient is not explained by the exemplar model, but is only accounted for by the prototype model. To check this, we entered the predictions of both the exemplar model and the feature-based prototype model[7] as predictors in a regression analysis with the observed typicality as criterion. In this way, we can investigate the differential contribution of the exemplar and prototype model in the prediction of typicality. The results of these analyses are shown in Table 1.

Table 1

R-squared and b-coefficients of the feature-based prototype predictor and the exemplar predictor for the 10 complex concepts. Note that for the exemplar predictor, the dimensionality was set at 5

Concept	R-squared	prototype	exemplar
Outdoor sports	0.51	-0.005	9.351**
Indoor sports	0.36	0.080	5.814**
Rock music	0.61	0.141*	6.911**
Classical music	0.39	0.021	4.326**
Transport people	0.70	0.080*	4.194**
Transport goods	0.70	-0.010	8.684**
Summer clothes	0.53	0.089	3.479**
Winter clothes	0.57	0.067*	4.088**
War weapons	0.95	0.052	4.183**
Sport weapons	0.53	0.163	7.549**

* $p < .05$ ** $p < .01$

7. We did not include the spatial prototype model in these analyses due to problems of colinearity.

Table 1 shows that in the regression analyses the exemplar model is clearly the dominant predictor. In all complex concepts, the exemplar model contributes significantly (at level .01) to the prediction of typicality, while the feature-based prototype model does not contribute significantly at level .01 and only in 3 of the 10 concepts, at level .05. These results strongly suggest that there is little or no variance in the observed typicality ratings explained by the feature-based prototype model that is not accounted for by the exemplar model.

Alternative prototype approaches

The previous section presents strong evidence that an exemplar approach is a more than viable candidate to account for complex concepts. However, while the two prototype models that were applied – the reweighted feature-based approach and the average spatial prototype – are the most common and dominant implementations of the prototype approach, they are not the only alternatives within the prototype approach that deserve consideration, particularly in the context of natural language concepts. To challenge the exemplar approach further, a number of alternatives to the dominant implementations of a prototype can be considered. In the present section, we will give a description of three – related – prototype models that subsequently will be tested against the exemplar model.

A median prototype instead of an average prototype

A first variant within the prototype approach differs slightly in its translation of a prototype to a spatial framework. Whereas the spatial prototype model applied earlier, proposes that the prototype is the average of a category, one can also consider a median as possible prototype. The most important difference between the average and the median in the present context is that a median is less sensitive to outliers. In effect, a median prototype will be more similar to the majority of the category, and less similar to potential outliers of the category.

Formally, the typicality according to the median prototype approach can be calculated by applying Equation 3, and use median values of a subset on each dimension to determine the location of the prototype.

A feature occurrence weight

In the feature-based prototype model, it is assumed that the prototype of a complex concept consists of a reweighting of the feature structure of the modified concept. In the previous section, the reweighting was implemented through feature importance ratings. Another way to arrive at appropriate fea-

ture weights is by considering the occurrence of a particular feature in the members of the complex concept. Instead of using ratings of how important a feature is for a complex concept, the occurrence of a feature in the extension of a complex concept indicates to what extent members of the complex concept have the feature.

Formally, a measure of typicality can be derived for the feature-based prototype on the basis of feature occurrence by applying Equation 1, substituting the feature importance ratings by feature occurrence counts. This model will be referred to as the feature occurrence prototype model.

A feature diagnosticity weight

Instead of taking into account solely the occurrence of a feature in the extension of a modified concept, it might also be of importance to consider whether the feature occurs in the members of other (related) concepts (e.g., Costello & Keane, 2001; Rosch, 1978). A useful notion capturing this idea is *diagnosticity*. Diagnostic features are features that occur frequently in members and only rarely in non-members. In the present prototype variant, we assume that diagnostic features are weighted more heavily in the construction of the prototype (that is, the reweighted feature structure).

Formally, a feature-based account of typicality can be derived by applying Equation 1 and substituting the feature importance ratings with feature diagnosticity. The diagnosticity of a feature for a complex concept, $D(x|A)$ can be determined by (see, e.g., Costello, 2000):

$$D(x|A) = \frac{\sum_{j \in A} j_x}{|A| + \sum_{i \in K} i_x} \tag{5}$$

where j_x is 1 if feature x applies to exemplar j, $|A|$ indicates the number of members in C, and K refers to the set of exemplars that is not in C. As can be easily seen, a feature has a diagnosticity value of 1 when it is present in all members of the complex concept A and not in non-members. As soon as a feature is not applicable to a particular member, or applicable to non-members, the diagnosticity measure decreases. The feature diagnosticity measure can then be used as a weight for each feature in the calculation of typicality scores following Equation 1. This model is referred to as the feature diagnosticity prototype model[8].

8. The feature occurrence prototype model and the feature diagnosticity prototype model are constructed on the basis of individual instances, and hence have characteristics of exemplar models. Here they are considered prototype models because they imply a reweighting of the prototype (that is, the feature structure) of the unmodified concept.

Comparison with the exemplar model

The three models were compared to the exemplar based model in their account of the typicality gradient of the 10 complex concepts. A summary of the results is presented in Table 2. First, we present the average correlation across complex concepts (column 4). For the spatial models, the dimensionality was fixed at 8 for these analyses. The results confirm the conclusion of the previous section: The exemplar model in general provides a better fit than the prototype models. Of the 5 prototype models, the spatial prototype models (both the average and the median prototype) provide the best prototype accounts (respectively $r = .71$ and $r = .70$), followed by the feature diagnosticity prototype model ($r = .62$), but still are clearly outperformed by the exemplar model[9].

Table 2

Comparison of the exemplar model to the five prototype models. The third column presents the average correlation (across categories) of the six models (for the spatial models, the Dimensionality is fixed at 8). Columns 4 to 8 present the number of complex concepts (out of 10) for which a particular model provides the best account of the typicality gradient, as a function of dimensionality of the spatial representation. Note that the dimensionality only applies to the spatial models

	Models	average correlation	dimensionalities				
			4d	5d	6d	7d	8d
Spatial models	Exemplar model	.75	6	7	7	7	7
	Average prototype	.71	0	0	0	0	0
	Median prototype	.70	0	0	1	1	1
Feature-based prototype models	Feature importance	.52	1	0	0	0	0
	Feature occurrence	.48	0	0	0	0	0
	Feature diagnosticity	.65	3	3	2	2	2

Columns 4 to 8 in Table 2 give more details on the model comparisons, presenting the number of complex concepts for which a particular model provides the best typicality account, as a function of the dimensionality of the underlying spatial representation. The results confirm the previous findings: the exemplar model quite consistently provides a better account of typicality in the complex concepts. In only three concepts, one of the five prototype models outperforms the exemplar model. The median prototype model performs best in 'musical instruments used in classical music' from Dimension-

9. The reported pattern of results is identical in all dimensionalities under consideration, that is, while the data fit of the spatial models tends to improve in higher dimensionalities, the rank order of all six models is identical in terms of fit across dimensionalities.

ality 6 onwards, and the feature diagnosticity model is successful in 'vehicles used for the transport of people' and 'winter clothes'.

In sum, even when comparing the exemplar model to five variants of the prototype approach, it remains the most competitive model by far. While we find some evidence for a prototype approach in three complex concepts, none of the prototype models succeeds in providing a satisfying account of typicality for all complex concepts. These results confirm that the exemplar approach to complex concepts is clearly more successful than any single prototype model under consideration.

Number of instantiated exemplars

A final issue relates to the number of exemplars that are instantiated when building the exemplar model's representation for the complex concepts. Looking at the correlation observed and predicted typicality as a function of the number of exemplars included in the representation for Dimensionality 5 – as presented in Figure 2 – it becomes clear that, in general, there seems to be an optimal number of instantiated members for each category, generally somewhere between 5 and 10 instances[10]. For most categories, the predictive correlation clearly decreases when the number of instantiations grows above ten.

For the categories 'vehicles for transporting people' and to a lesser degree 'musical instruments used in classical music', results are less clear, the former showing a decrease in fit when including 2 to 8 instances, and from there onwards a steadily increasing fit. The latter does not seem to have a clear optimal number of instances – the optimal number could be anywhere between 5 and 15.

Discussion

In research concerning artificial category learning and in research involving natural language concepts, it is found across an impressively large array of conditions that exemplar models in general provide a better description of human categorisation than abstractionist prototype models. Nosofsky (1992) and Vanpaemel and Storms (2010) provide overviews of this computational research. Nosofsky (1992) summarises 36 studies, of which only 6 favour a prototype representation over an exemplar representation. Vanpaemel and Storms (2010) review 30 studies, all by Nosofsky and his collaborators, only three of which provide evidence for prototype representations. Studies comparing exemplar and prototype representations in natural language categories

10. The same pattern was observed in the other Dimensionalities.

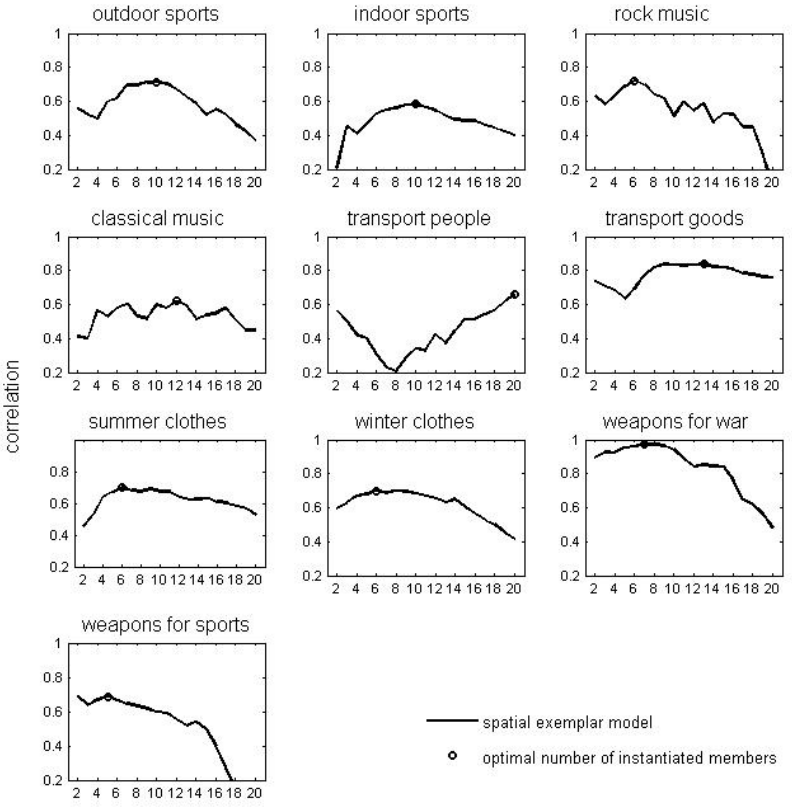


Figure 2

Correlations between observed and predicted typicality in function of the number of exemplars included in the representation for the exemplar predictor (solid line). The circle points to the optimal number of instantiated elements. The dimensionality is fixed at 5

are few and far between, but also support the same conclusion: exemplar representations are found to be superior to prototype representations in their account of category dependent judgments such as typicality (Storms, De Boeck, & Ruts, 2000; Voorspoels et al., 2008).

Despite the clear dominance of exemplar models of representation, there has been little effort to implement the exemplar view in a crucial characteristic of the human conceptual apparatus: productivity and creativity, that is to say, the ability to flexibly combine concepts to more complex wholes (e.g., Fodor & Lepore, 2002; Murphy, 2002). The phenomenon of combining simple concepts to complex wholes has been largely analysed from a prototype

perspective, giving only a secondary role to exemplar information, for example in terms of extensional feedback (for exceptions, see Costello, 2000; Gray & Smith, 1995). In the present study, we implement and evaluate a straightforward spatial exemplar model, in which the representation of a complex concept is made up by a subset of exemplars of the unmodified category. These exemplars are instantiated in the representation construction process of the complex concept. According to the spatial exemplar model, typicality of an exemplar for a complex concept was defined as the similarity of an exemplar to the subset of instantiated exemplars.

To evaluate whether the exemplar approach is viable in the context of concept combination, we tested it against the traditional prototype approach to conceptual combination. In order to provide an optimal challenge, we compared the exemplar model to five different and plausible implementations of the prototype approach, derived from two dominant views. On the one hand, we considered that conceptual combination essentially involves the reweighting of the feature structure of the concept that is modified (e.g., Murphy, 1990; Smith et al., 1988), and examined three different ways two arrive at the appropriate weights. On the other hand, we implemented the dominant idea of a prototype as a central tendency of a category (e.g., Minda & Smith, 2010), and tested both an average prototype and a median prototype.

The results clearly favoured the exemplar model. In comparison with the five prototype model variants, the exemplar model achieved the best account of typicality on average (across categories), and provided the best data fit in 7 out of 10 complex concepts. For three complex concepts, one of the prototype models performed better than the exemplar model, but none of the prototype model variants was able to provide an account of the observed typicality equally consistent across categories. Moreover, regression analyses including both the exemplar and prototype model predictions demonstrated that only a small proportion of the variance in the observed typicality ratings was uniquely accounted for by the prototype model. Taken together, our results strongly suggest that an exemplar approach can indeed be successfully implemented in the study of complex concepts and that exemplar information plays a role more fundamental than is currently acknowledged in the traditional theories of conceptual combination. This consequently underlines the importance of notions such as extensional feedback.

Obviously, the evidence we presented for the viability of the exemplar approach in conceptual combination crucially depends on the nature of the complex concepts presented to the participants. In particular, the extent to which a combination of words refers to a complex concept that requires the construction of a representation, needs careful consideration. A combination of words, such as used in the present study, does not necessarily refer to a combination of concepts, but can simply serve as a description for a readily

available concept in semantic memory, for example in case of highly familiar and frequently used, lexicalised conceptual combinations. At face value, the complex concepts in the present study do require the combination of several concepts (e.g., the concepts of 'sports' and 'weapons' to arrive at 'weapons used in sports'). Moreover, while the complex concepts undoubtedly differ in novelty, the dominance of the exemplar model is consistent across categories. For example, the concepts 'summer clothes' and 'indoor sports' are perhaps more familiar and the concepts 'musical instruments used in rock music' or 'weapons used for sports' are more novel. But, for all four of these complex concepts the exemplar model provided the best account of the typicality gradient. While there is no straightforward criterion to decide which concepts are "simple", and which are combinations of different concepts based on the words in a language (see e.g., Barsalou, 1991), the consistence across concepts suggests that the present conclusion does not depend on the familiarity of the complex concepts.

Relation to earlier findings and implications

Our findings regarding exemplar representations in complex concepts are remarkably similar to earlier findings in basic natural language concepts (Storms et al., 2000), in which it was established that an exemplar representation of a category, including 5 to 10 instances, seems to result in the best prediction of typicality. This striking parallel suggests an intriguing interpretation. It has already been suggested that much of the structure of concepts *as they are used* is constructed when it is needed, "on the fly", from a large body of associative knowledge, and that invariant concepts that are stored in long term memory are an "analytic fiction" (e.g., Barsalou, 1987; Barsalou, 1991; Barsalou, 1999; Prinz, 2002). More specifically regarding the graded structure of concepts, there is empirical evidence that suggests that this structure is instable and flexible. In particular, typicality judgments can be instable both within and between participants (Barsalou, 1987) and situational and sentential context have been found to affect typicality judgments (Roth & Shoben, 1983). Whereas one would suspect that only less-established, new concepts – such as complex concepts resulting from combinations – require a representation construction process, these findings suggest that the same holds for well-established concepts.

Looking at our results in the light of these remarks, the similar patterns found in complex concepts and simple concepts suggest that the cognitive process underlying evaluations of typicality is not fundamentally different: judgments of typicality are based on a representation that is created consisting of the instantiation of 5 to 10 members of the category, both for complex concepts and well-established categories. An obvious strength of the instantia-

tion-based exemplar approach evaluated in this study is that it is compatible with models used in simple concept research. As such, the model could be a step towards a more unified theory of concepts, covering a broader range of phenomena.

In this study we left open the essential question of how the right members of the complex concept are instantiated. We do not in any way claim that the categorisation data we used in the instantiation process has any explanatory value nor do we at the moment have a viable alternative. While the instantiation-based exemplar model performed well, the crucial question for this model is: how can we construct novel, unfamiliar complex concept representations if we have no remembered instances to call to mind (Hampton, 1997; Rips, 1995). For now, this obviously is an important shortcoming of the exemplar model as presented here, and a notable disadvantage in comparison with an earlier attempt at implementing exemplar models in conceptual combination (Costello, 2000). On the other hand, the model is not restricted to using categorisation data. Other variables – such as familiarity or association strength – with more explanatory strength can be implemented in the same way. In this sense, the instantiation-based exemplar model allows the explicit study of such variables, and could perhaps be a valuable tool in the systematic study of conceptual combination.

References

Barsalou, L.W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge: Cambridge University Press.

Barsalou L.W. (1991). Deriving categories to achieve goals. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory, vol. 27* (pp. 1-64). San Diego: Academic press.

Barsalou, L.W. (1999). Perceptual symbols systems. *Behavioral &Brain Sciences, 22*, 577-660.

Brooks, L.W. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization* (pp. 169-211). Hillsdale, NJ: Erlbaum.

Busemeyer, J.R., Dewey, G.I., & Medin, D.L. (1984). Evaluation of exemplar-based categorization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 638-648.

Costello, F.J. (2000). An exemplar model of classification in single and combined categories. In L.R. Gleitman & A.K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 95-100). Hillsdale, NJ: Erlbaum.

Costello, F.J. (2001). Testing a computational model of categorisation and category combination: Identifying disease categories and new disease combinations. In

J.D. Moore & K. Stenning (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 238-243). London: Lawrence Erlbaum Associates.

Costello F.J., & Keane, M.T. (2001). Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 255-271.

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M.J., Voorspoels, W., & Storms, G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods, 40*, 1030-1048.

Devereux, B., & Costello, F. (2007). Learning to interpret novel noun-noun compounds: Evidence from a category learning experiment. *Proceedings of the workshop on cognitive aspects of computational language acquisition* (pp. 89-96). New York: Lawrence Erlbaum Associates.

De Wilde, E., Vanoverberghe, V., Storms, G., & De Boeck, P. (2003). The instantiation principle re-evaluated. *Memory, 11*, 533-548.

Estes, Z.C., & Glucksberg, S. (2000). Interactive property attribution in concept combination. *Memory & Cognition, 28*, 28-34.

Fodor, J.A., & Lepore, E. (2002). *The compositionality papers*. Oxford: Oxford University Press.

Gagné, C.L., & Shoben, E.J. (1997). Influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 71-87.

Gray, K.C., & Smith, E.E. (1995). The role of instance retrieval in understanding complex concepts. *Memory and Cognition, 6*, 665-674.

Hampton, J.A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 12-32.

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

Hampton, J.A. (1997). Conceptual Combination. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 135-162). Hove: Psychology Press.

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

Helft, M. (2008). Ad accord for Yahoo and Google. *New York Times*. http://www.nytimes.com/2008/06/13/technology/13yahoo.html?_r=1&ref=todayspaper&oref=slogin

Levi, J.N. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

Maguire, P., Devereux, B., Costello, F., & Cater, A. (2007). A reanalysis of the CARIN theory of conceptual combination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 811-821.

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

Medin, D.L., & Shoben, E.J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20*, 158-190.

Minda, J.P., & Smith, J.D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 775-799.

Minda, J.P., & Smith, J.D. (2010). Prototype models of categorization: Basic formulation, predictions, and limitations. In E. Pothos & A. Wills (Eds.), *Formal approaches in categorization* (pp. 117-162). Cambridge University Press: Cambridge, UK.

Murphy, G.L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language, 29*, 259-288.

Murphy, G.L. (2002). *The big book of concepts*. Cambridge, MA: MIT press.

Nosofsky, R.M. (1992). Exemplars, prototypes, and similarity rules. In A.F. Healy & S.M. Kosslyn (Eds.), *Essays in honor of William K. Estes, vol. 1: From learning theory to connectionist theory; vol. 2: From learning processes to cognitive processes* (pp. 149-167). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Osherson, D.N., & Smith, E.E. (1981). On the adequacy of prototype theory of as a theory of concepts. *Cognition, 9*, 35-58.

Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353-363.

Prinz, J.J. (2002). *Furnishing the mind: Concepts and their conceptual basis*. Cambridge, MA: MIT Press.

Rips, L.J. (1995). The current status of research on concept combination. *Mind and Language, 10*, 72-104.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization* (pp. 28-48). Hillsdale, NJ: Erlbaum.

Roth, E.H., & Shoben, E.J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15*, 346-378.

Smith, E.E., & Osherson, D.N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8*, 337-361.

Smith, E.E., Osherson, D.N., Rips, L.J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science, 12*, 485-527.

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

Vanpaemel, W., & Storms, G. (2010). Abstraction and model evaluation in category learning. *Behavior Research Methods, 42*, 421-437.

Verheyen, S., Ameel, E., & Storms, G. (2007). Determining the dimensionality in spatial representations of semantic concepts. *Behavior Research Methods, 39*, 427-438.

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Exemplars and prototypes in natural language concepts: A typicality based evaluation. *Psychonomic Bulletin & Review, 15*, 630-637.

Wisniewski, E.J. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language, 38*, 177-202.

Wisniewski, E.J., & Middleton, E.L. (2002). Of bucket bowls and coffee cup bowls: Spatial alignment in conceptual combination. *Journal of Memory and Language, 46*, 1-23.

Wisniewski, E.J., & Murphy, G.L. (2005). Frequency of relation type as a determinant of conceptual combination: A reanalysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 169-174.

Younger, B.A., & Cohen, L.B. (1983). Infant perception of correlations among attributes. *Child Development, 54*, 858-867.

Received September 12, 2011
Revision received January 18, 2012
Accepted March 7, 2012