The generalized polymorphous concept account of graded structure in abstract categories

**Steven Verheyen · Loes Stukken · Simon De Deyne ·
Matthew J. Dry · Gert Storms**

Abstract Abstract categories present with graded structure. The extent to which feature commonality between exemplars and category provides a satisfying account of this graded structure varies from one abstract category to the other (Hampton, 1981). We investigate whether the incorporation of features that exemplars share with external categories yields an improved account of abstract categories' graded structures. In doing so, we follow the suggestion that abstract categories are relational in nature (Goldstone, 1996; Wiemer-Hastings & Xu, 2005). The generalized polymorphous concept model, which incorporates both types of features, is found to improve the account of typicality and category membership in three of seven studied abstract categories. These three categories are found to be the most abstract, suggesting that it is appropriate to think of abstract categories as varying along a continuum of abstractness/interrelatedness rather than as a distinct type of category altogether.

Keywords Abstract categories · Relational categories · Typicality · Category membership · Feature representation

S. Verheyen (✉) · L. Stukken · S. De Deyne · M. J. Dry · G. Storms
Department of Psychology, University of Leuven,
Tiensestraat 102, Bus 3721,
BE-3000 Leuven, Belgium
e-mail: steven.verheyen@psy.kuleuven.be
URL: http://ppw.kuleuven.be/concat/

M. J. Dry
Discipline of Pharmacology, University of Adelaide,
Adelaide, SA 5005, Australia

Introduction

Well-defined categories (e.g., *odd numbers*, *plane geometry figures*, Armstrong, Gleitman, & Gleitman, 1983), ad hoc categories (e.g., *Things that could fall on your head*, *Ways to make friends*, Barsalou, 1983), abstract categories (e.g., *crimes*, *sciences*, Hampton, 1981), and concrete categories (e.g., *birds*, *vehicles*, Rosch, 1973) are alike in that they all present with graded structure. That is, some of their exemplars are consistently judged to be more representative for the category than others. For example, a *sparrow* is generally considered to be a more representative member of the category *birds* than is a *penguin*.

Graded category structure also becomes apparent in category verification tasks in which representative exemplars tend to be endorsed more quickly than unrepresentative exemplars. Rosch (1973) was among the first to demonstrate the speed-gradedness relationship in concrete categories. Armstrong et al. (1983) and Hough and Pierce (1989) demonstrated that it is also present in well-defined and ad hoc categories. In addition, representative exemplars are generally among the first to be generated in response to the category label and, across participants, they are generated more often than are less representative category members (for early demonstrations in concrete and ad hoc categories, see Mervis, Catlin, & Rosch, 1976, and Barsalou, 1985, respectively).

Category types

Much of our understanding of the nature of categories arises from the study of concrete categories, such as *birds* or *vehicles*. However, Medin, Lynch, and Solomon (2000) argued that extending the scope of studies to include other types of categories is important because it tests the

generality of prevailing theories and has the potential to highlight differences between the underlying representations of the different category types. With categories considered to be the building blocks of cognition (Pinker, 1997), the question of which categories constitute truly distinct kinds (and which ones do not) becomes one of central importance in cognitive science.

Indeed, the presence of graded structure in a varied range of category types need not necessarily indicate that these category types are all represented in the same manner. There is evidence to suggest that some of the fundamental characteristics of concrete categories are not present in other category types. For example, it has been demonstrated that for many concrete categories, the more features an exemplar shares with the category as a whole, the more representative the exemplar tends to be (Hampton, 1979; Rosch & Mervis, 1975). Since *penguins* do not have features<*can fly*>and<*build nests in trees*>in common with most other members of the category *birds*, they are considered to be unrepresentative category members. In other words, it would appear that for concrete categories, the more similar an exemplar is to the category as a whole, the more representative it will be.

In contrast with this assertion, the internal structure of many ad hoc and well-defined categories does not adhere to this account. Barsalou (1985) demonstrated that participants' estimates of the number of times they had previously encountered an item as a category member and the item's ability to fulfill the goal served by its ad hoc category provided a better account of the item's judged representativeness than did a similarity-based measure. Furthermore, Larochelle, Richard, and Soulières (2000) ascribed well-defined categories' apparent graded structures in category verification to the exemplars' familiarity and category dominance. They demonstrated that when the influence of these variables was controlled for, reliable categorization time differences disappeared.

Studies addressing the internal structure of abstract categories are less common, but nevertheless also seem to suggest that their representations differ from those of concrete categories in a number of important ways. Medin et al. (2000) discerned two (interrelated) criteria that can be used to argue for the different nature of concrete and abstract categories. These criteria pertain to the processing differences and the structural differences between both types of categories.[1] Lakoff and Johnson (1980) introduced

_____

[1] Medin et al. (2000) included content-laden principles as a third criterion, but we will not discuss it here since, as far as we know, no principled attempts have been made to distinguish concrete and abstract categories on the basis of the causal explanatory frameworks they invoke.

the idea that abstract entities are processed through reference to other, more concrete, entities. For example, *anger* can be understood by referring to water that comes to a boil. According to this idea, abstract categories are processed differently from concrete ones, in that the latter ones do not require metaphorical mapping. Structural differences support such claims, in that the features that are generated in response to abstract entities are less specific than the features generated for concrete ones (Wiemer-Hastings & Xu, 2005). Abstract categories also invoke more relational features than concrete categories do, but do not activate as many entity features (Barsalou & Wiemer-Hastings, 2005; Wiemer-Hastings & Xu, 2005). Wiemer-Hastings and Xu took this to mean that abstract categories are relational categories (see Goldstone, 1996, for a similar suggestion). Relational categories have a relatively weak internal structure, but strong links to external categories (Gentner, 1981; Markman & Stilwell, 2001). Thus, there is both processing and structural evidence to suggest that a key difference between abstract and concrete categories lies in the greater contribution of external categories to the representation of the former as opposed to that of the latter.

This is of particular interest for a study by Hampton (1981) on the gradedness of abstract categories. It involved the polymorphous concept model—a measure of the number of features shared by a category and its constituent exemplars. Hampton had participants generate characteristic features for eight abstract categories. Subsequently, he had a different group of participants judge whether these features were applicable to the categories' exemplars. The degree to which category and exemplars shared characteristic features had proven to account for gradedness in various concrete categories (Hampton, 1979), but didn't fare as well in the abstract categories studied by Hampton (1981). Although feature commonality between exemplars and category correlated with gradedness in all eight studied categories, the relationship was much less pronounced for three of them. This might be the case because in abstract categories that are partly represented in terms of other categories, it might not suffice to take into consideration only the features an item shares with the target category (i.e., target features). Features that the item shares with other categories (i.e., external features) might also be of importance in determining how representative an item is for the target category. This rationale underlies the generalization by Dry and Storms (2010) of the polymorphous concept model (Hampton, 1979, 1981). Dry and Storms demonstrated how the model can be generalized to include both target and external features. Allowing external features to influence the assessment of exemplars' similarity to the category might improve our account of abstract categories' graded structures. Perhaps the apparent difference between the origins of the graded structure of the concrete categories

in Hampton (1979) and of some of the abstract categories in Hampton (1981) can be made to disappear by including external features. If the generalized polymorphous concept model were to provide a satisfying account of all abstract categories' graded structures, this would suggest that exemplar-category similarity (expressed as a combination of target and external features) is the major determinant of gradedness in both concrete and abstract categories. To see this, consider the following exposition of Dry and Storms' generalization of the polymorphous concept model.

The generalized polymorphous concept model

The polymorphous concept model (Hampton, 1979) assumes that features are the representational units of semantic concepts and that the more of these features a particular exemplar and its category share, the more representative the exemplar is considered to be. If both the category and the exemplars are represented by feature vectors v of length k in which the absence of a characteristic feature is signified by 0 and its presence by a positive integer (signaling the feature's importance, its salience, or the agreement that exists about the former), then the representativeness of an exemplar i with respect to category A can be formalized as:

$$PC(i,A) = \sum_k v_{ik} v_{Ak}. \qquad (1)$$

Because of the multiplicative nature of Eq. (1), only those features that are shared by i and A contribute to exemplar i's representativeness. Earlier, we already gave the example of the *penguin*, an animal considered to be atypical among *birds*, since it cannot fly and does not build its nest in trees, like other birds do. Penguins, however, also<*swim*>and<*have flippers*>. These are features that the penguin shares with exemplars of the category of *fish* and might plausibly have an impact on its perceived representativeness as a bird. Dry and Storms (2010) have made the relevance of external features for typicality and membership judgments more salient by casting the polymorphous concept model in terms of exemplar-category similarity. From the work of Tversky (1977), it has become apparent that both common and distinctive features contribute to similarity. Following Tversky's contrast model, the similarity s_{iA} between an exemplar i and a category A can be expressed as a weighted combination of the features that i and A share ($v_i \cap v_A$), the features that are distinct to i ($v_i - v_A$), and the features that are distinct to A ($v_A - v_i$):

$$s_{iA} = \alpha f(v_i \cap v_A) - \beta f(v_i - v_A) - \gamma f(v_A - v_i), \qquad (2)$$

where f is a monotonic function, and $0 \leq \alpha, \beta, \gamma \leq 1$. From this, it can be seen that if $\alpha > \beta = \gamma = 0$, then similarity is

based purely on common features, and if $\beta, \gamma > \alpha = 0$, then similarity is based purely on distinctive features. Following this, the polymorphous concept model in Eq. 1 can be written in a generalized form as:

$$GPC(\theta, i, A) = \left[\theta \sum_k v_{ik} v_{Ak} \right]$$
$$- \left[(1-\theta) \sum_k v_{ik}(1 - v_{Ak}) \right]$$
$$- \left[(1-\theta) \sum_k (1 - v_{ik}) v_{Ak} \right]. \qquad (3)$$

In Eq. 3, a single parameter θ (ranging from 0 to 1)—rather than three separate parameters (α, β, and γ)—signals the contributions of shared and distinctive features. This parameter allows one to express the degree of emphasis given to common versus distinctive features using a single value (Navarro & Lee, 2004). Setting θ to low values emphasizes distinctive features, whereas setting θ to high values emphasizes common features.

Dry and Storms (2010) noted that the first and third terms in Eq. (3) are collinear. Each category A will be characterized by a fixed number of features. In that case, the number of features shared by A and i, plus the number of features possessed by A and not by i, will equal this fixed number. Hence, the first and third term in Eq. (3) will sum to a constant, and one of them can be dropped from the equation without loss of information. Dry and Storms reformulated Eq. (3) by dropping the third term so that the expression for the generalized polymorphous concept becomes:

$$GPC(\theta, i, A) = \left[\theta \sum_k v_{ik} v_{Ak} \right]$$
$$- \left[(1-\theta) \sum_k v_{ik}(1 - v_{Ak}) \right]. \qquad (4)$$

The θ parameter retains its interpretation in Eq. (4). It indicates the relative contribution of features that exemplar i shares with category A and of features that exemplar i does not share with category A. The latter ones can, in principle, be truly distinctive features. This is the case when they are unique to the exemplar. More often, however, these will be features that are deemed characteristic of a category other than the target. It is then appropriate to term them external features, since they signal overlap with an external category. In all of our studies, the distinctive features will be features that are characteristic of an external category. It is in this sense that we will therefore interpret the second term of Eq. (4).

In the following sections, we will determine the value of θ in (4) that provides for an optimal account of abstract

categories' graded structures. If these categories are indeed relational in nature, allowing θ to differ from 1 should improve the correlation with measures of gradedness, indicating that features that the exemplars share with external categories (i.e., external features) contribute to their representativeness.

Study 1

The study by Hampton (1981) is so far the only one addressing the graded structure of abstract categories. Therefore, we feel it important to start by replicating his finding that feature commonality between exemplars and category does not accord well with all abstract categories' graded structures. Rather, the extent to which the original polymorphous concept model accounts for their internal structure is expected to differ from one abstract category to the other. In the following section, we will describe the materials necessary to allow a replication of Hampton (1981), along with the procedures employed to gather them.[2] Unlike Hampton (1981), who used only membership judgments, we obtain both membership and typicality judgments as measures of category gradedness. Although both measures generally tend to correlate very strongly, some authors have pointed toward subtle differences between them (e.g., Osherson & Smith, 1997; Rips, 1989). These differences might become more pronounced in a study that pertains to the influence of external feature information on category gradedness. It is not unlikely that the question of category membership invokes the possibility that the items might belong to other categories than the target to a greater extent than the question of typicality does. The assumption of membership might be implicit in the question of how typical an item is of a target category. The contribution of external features might therefore be more pronounced for judgments of category membership than for judgments of typicality.

Exemplar generation

All seven of the categories adhere to Hampton's (1981) definition of an abstract category in that they have as referents things that are not physical, concrete objects. As part of a large exemplar generation study that included 30 categories of a varied nature, 80 first-year psychology students of the University of Leuven produced eight exemplars for the categories *art forms*, *crimes*, *diseases*, *emotions*, *media*, *sciences*, and *virtues*. A tally was kept of the number of times a particular exemplar was generated in

response to the category label. This tally informed the selection of exemplars. Because previous work on concrete and ad hoc categories (Barsalou, 1985; Mervis et al., 1976) has established a strong relationship between generation frequency and representativeness, we hoped to obtain exemplars that differed considerably in representativeness by selecting 15 exemplars that spanned the range of generation frequencies for a particular abstract category. The Appendix holds the selected exemplars per category, in descending order of generation frequency.

Typicality judgments

A group of 30 first-year psychology students of the University of Leuven provided typicality judgments for the 15 selected exemplars of each of the seven abstract categories. They received a booklet containing seven pages with, on each page, the 15 exemplars of one of the seven categories. The participants were asked to indicate for every item on the page how good an example it was of the category mentioned on top of the page. They were required to provide their answer by indicating a value on a scale ranging from 1 (*a very bad example*) to 20 (*an excellent example*). If they did not know one of the presented exemplars, they were asked to indicate this by drawing a circle around it. The same instructions were used by De Deyne et al. (2008) to obtain typicality judgments for exemplars of concrete categories. Every participant rated the typicality of all the exemplars of every category. There were two different presentation orders for the categories and three different presentation orders for the exemplars within a category. This resulted in six different booklets that were each completed by five participants.

To confirm the validity of the exemplar selection procedure described previously, the Pearson correlation between the typicality judgments and the log-transformed generation frequencies was calculated for each category. The correlations were $r = .85$ for *art forms*, .63 for *crimes*, .57 for *diseases*, .89 for *emotions*, .90 for *media*, .71 for *sciences*, and .60 for *virtues*. All of these correlations were significant at the $p = .05$ level (one-tailed t). These results establish that the commonly found relationship between typicality and generation frequency also holds in abstract categories.

Category membership judgments

A group of 30 first-year psychology students of the University of Leuven provided category membership judgments for the 15 selected exemplars of each of the seven abstract categories. Aside from the instructions, the procedure employed to obtain these category membership judgments was identical to that used to obtain typicality judgments. The instructions were

taken from Estes (2004). They asked participants to indicate the exemplars' degree of category membership on a scale ranging from 0 (*not at all a member*) to 10 (*completely a member*).

The obtained membership judgments also proved to correlate significantly at the $p = .05$ level (one-tailed t) with the log-transformed generation frequencies. The correlation was established at .72 for *art forms*, .69 for *crimes*, .53 for *diseases*, .82 for *emotions*, .87 for *media*, .62 for *sciences*, and .58 for *virtues*. That judged category membership, like typicality, consistently shows a relationship with generation frequency is not surprising in light of the strong correlations that were obtained between the judgments of category membership and typicality. The correlation measured .84 for *art forms*, .97 for *crimes*, .88 for *diseases*, .92 for *emotions*, .98 for *media*, .97 for *sciences*, and .93 for *virtues*.

Feature generation

Thirty University of Leuven students completed a task intended to elicit the features associated with each of the categories. They answered three questions for each of the category labels: (a) Which features do you feel are important for this category? (b) Which features have to be present for something to be considered a member of this category? (c) Which features determine that something is a better example of this category than something else? Every participant answered these questions for each of the seven abstract categories. Each participant was presented with a different (random) order of categories. Participants could generate as many features as they felt necessary. Across participants and questions, 46 different features were generated for *art forms*, 38 for *crimes*, 50 for *diseases*, 44 for *emotions*, 35 for *media*, 46 for *sciences*, and 40 for *virtues*.

Feature applicability

The exemplars and features that were generated in response to the category labels were combined in a feature x exemplar matrix. The $46 + 38 + 50 + 44 + 35 + 46 + 40 = 299$ features (in alphabetical order) made up the rows of the matrix, whereas the $7 \times 15 = 105$ exemplars (also in alphabetical order) made up the columns of the matrix. Hence, every feature–exemplar pair was represented by a single cell in the matrix. Five University of Leuven students indicated for each feature–exemplar pair whether the feature applied to the exemplar or not, by entering a 1 or a 0 in the corresponding matrix cell. Participants performed the task at home and could freely choose when they worked on it. They were given the choice to work on the task row-wise or column-wise, but they were asked not to pause until a row or column was finished.

Model analyses

The five feature x exemplar matrices were summed to form a single matrix. The resulting matrix provided the exemplar feature vectors v_i that enter into Eq. (4). The entries in v_i thus vary between 0 and 5, reflecting, for each feature, the number of students who indicated that it applies to item i. The category feature vector v_A in Eq. (4) holds ones for those features that were generated in response to the category label in the feature generation task, and zeros for those that were not. For each of the seven abstract categories, the original polymorphous concept measure was then computed and correlated with judged typicality and judged category membership. This amounts to fixing θ in Eq. (4) to 1 and thus taking into account only those features that the exemplars and the target category share in common. It is also equivalent to restricting the analyses to an applicability matrix that includes only the features that were generated in response to the target category.

The results for typicality resemble those of Hampton (1981) in that feature commonality between exemplars and category is found to account for the graded structure of some of the abstract categories, but not of all. The correlation between the original polymorphous concept measure and typicality was established at .74 for *art forms*, at .79 for *crimes*, at .78 for *emotions*, and at .90 for *media*. These four correlations were found to be significant at the .01 level (one-tailed t). The correlations for *diseases* and *sciences* were somewhat less pronounced with $r = .58$ and $r = .48$, respectively, but still reached significance at the .05 level. The .09 correlation for *virtues* was not significant, however.

A similar pattern of results was found when the relationship between category membership and the original polymorphous concept measure was established. Including only target feature information led to a replication of the Hampton (1981) findings. The graded structure of some abstract categories was well accounted for by the original polymorphous concept measure. Its correlation with category membership was established at .61 for *art forms*, .87 for *crimes*, .67 for *emotions*, and .89 for *media*. These correlations are all significant at the .01 level, according to one-tailed ts. The correlation for *diseases* was established at .57 ($p < .05$). The correlation for *sciences* proved to be marginally significant ($r = .44$, $p = .05$), whereas the correlation for *virtues* was found to be not significant ($r = .04$).

There appears to be considerable correspondence between the results for typicality and category membership. When typicality is well accounted for by the target features measure, category membership appears to be well accounted for also. When the target features measure yields an unsatisfactory account of typicality, the account of category membership is unsatisfactory as well. The original polymorphous concept model appears to fare well in the

abstract categories of *art forms*, *crimes*, *emotions*, and *media*. The categories of *sciences* and *virtues* constitute abstract categories for which the model does not fare so well. Depending on what one considers a satisfying correlation, the category of *diseases* can be considered to belong to the former or the latter class.

Study 2

The aim of the second study was to establish whether features that are characteristic of another category than the target contribute to its internal structure. The contribution of external features would testify to the relational nature of the abstract category involved. In the context of concrete relational categories, the external features generally originate from contrast categories (Dry & Storms, 2010; Verheyen, De Deyne, Dry, & Storms, 2010). *Contrast categories* are defined as mutually exclusive terms that are organized under an inclusive covering term (Goldstone, 1996; Martin & Billman, 1994). In our exposition of the generalized polymorphous concept model, we gave the example of the contrasting categories *birds* and *fish*, which are subordinate to the category of *animals*. Abstract categories, however, do not appear to be hierarchically organized (Crutch & Warrington, 2005; Hampton, 1981). As a consequence, they do not lend themselves as easily to categorization as concrete categories do. This probably explains why our attempt to verify whether any of our seven abstract categories would emerge as a contrast category for another failed when we employed the common procedure for eliciting contrasts (Malt & Johnson, 1992). We asked participants to imagine that they had heard a description of an entity and had ventured a first guess at its identity. The participants were further asked to imagine that this identification of the entity as a member of category *X* (one of our seven abstract categories) was incorrect. At that point, they were asked for a second guess at the entity's identity. None of the studied abstract categories was consistently provided as a response to this question. In fact, no consistent responses emerged, whatsoever.

We therefore adopted an alternative means of identifying categories that might exert an influence on a target category's internal structure. The procedure followed a suggestion by Verbeemen, Vanoverberghe, Storms, and Ruts (2001) and comprised the identification of the category (from a given set) that was regarded most similar to the target category. A minimum of similarity between two categories is required for there to be (an effect of) external features. If two categories pertain to completely different domains, it becomes unlikely that characteristic features of one of the categories will apply to exemplars of the other category.

Since abstract categories have as referents things that are not physical, concrete objects, perceptual input is presumably of little importance in the development of their representation. Rather, one will have to rely strongly on their use in language to acquire a full understanding of abstract concepts (Breedin, Saffran, & Coslett, 1994; Quine, 1960). We therefore derived the similarity between each of our seven abstract categories from the degree to which their exemplars are attested in similar syntactic environments (Pado & Lapata, 2007). Like the Hyperspace Analogue to Language model (HAL, Lund & Burgess, 1996) and Latent Semantic Analysis (LSA, Landauer & Dumais, 1997), this approach assumes that the meaning of words can be derived from the context in which they occur. In the present study, the contexts are thought of as syntactic relations, rather than neighboring words in sentences (HAL) or documents (LSA). Wiemer-Hastings and Graesser (2000) have already shown that the use of syntactic context elements allows one to successfully predict similarities between abstract concepts. In the following section, we will elaborate on the procedures employed to identify a potential contrast for each of our seven target categories. This will then allow us to apply the generalized version of the polymorphous concept model to account for the typicality and category membership judgments that were obtained in Study 1.

Corpus analysis

A varied corpus of contemporary spoken and written Dutch, containing 171.793 lemmas, was parsed with Alpino, a wide-coverage computational analyzer of Dutch (Bouma, van Noord, & Malouf, 2001). The dependency relations that were thus obtained were used to calculate a measure of similarity between each pair of exemplars under study (using the cosine). To reduce the effect of word frequency, the count distribution for each exemplar was transformed using the point-wise mutual information formula (Church & Hanks, 1991).

To determine which categories are most similar to one another, the similarities between the exemplars of all pairs of categories were averaged and subjected to an additive tree clustering procedure. This procedure generates a connected graph in which every pair of nodes is connected through a unique path. This path comprises arcs that have nonnegative weights (i.e., lengths) associated with them. The sum of the weights of the arcs that make up the unique path between two nodes represents the dissimilarity between the nodes. The shorter the path between two nodes, the more similar they are considered to be (Sattath & Tversky, 1977). The advantage of using cluster analysis is that it yields a more reliable estimate of the closest pair, since the clustering results are determined by all pairwise

similarities together, rather than just the pairwise similarity with the maximal value (Verbeemen et al., 2001).[3]

The result of the clustering procedure is presented in Fig. 1. The additive tree representation contains seven terminal nodes (larger circles in normal typography), one for each of the seven abstract categories. It contains five internal nodes (smaller circles in bold typography).[4] The spatial layout of the nodes reflects the modeled dissimilarities (Lee, 1999; Shepard, 1980). The results showed that the category of *media* was closest to *art forms*. They also showed that the category of *emotions* was the one closest to both *crimes* and *diseases*. The category of *virtues* was found to be the one most similar to *emotions* and *media*. *Art forms* was found to be most similar to *sciences*. *Media* was found to be most similar to *virtues*. In the application of the generalized polymorphous concept model, the characteristic features of these categories will take the role of external features when they apply to the respective target categories' exemplars.

Model analyses

For each target category, the summed feature x exemplar matrix obtained in Study 1 was reduced to include only those features that were generated in response to the target and its potential contrast. The resulting matrices provided the exemplar feature vectors v_i that enter into Equation (4). The category feature vector v_A in Eq. (4) held ones for those features that were generated in response to the target category label and zeros for those that were generated in response to the contrast category. (Note that retaining all of the features in the matrix and setting the entries in the category feature vector that do not pertain to the target category to 0 would have the result that features that originate from another external category than the potential contrast could act as external features. This was the modus operandi in Study 3.) For each of the seven abstract categories, the generalized polymorphous concept model was then implemented by varying θ from 0 to 1 in increments of .0005. For every value of θ, the correlation with judged typicality and judged category membership was computed. The bold lines in Fig. 2 display for each category the correlation of the generalized polymorphous concept
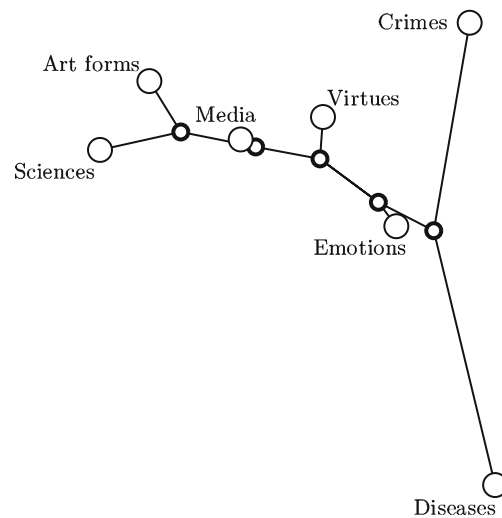

Fig. 1 Additive tree representation of the similarities between the seven abstract categories

with judged typicality across the entire range of θ. The dotted lines display the correlation with judged category membership.

The results for $\theta = 1$ are necessarily the same as the ones that were obtained in Study 1. A θ value of 1 in the generalized polymorphous concept model indicates that only the features that the exemplars share with their target category are taken in consideration. The resulting measure is thus not influenced by external features. The question then becomes whether we can improve on the model's account of typicality and category membership when external features are taken into consideration. To answer this question, we turn to Fig. 2 to establish the values of θ, which yield the optimal correlation with the measures of graded structure. As was the case in Study 1, the results seem to suggest that the abstract categories in question separate into two groups.

The θ values that yield an optimal correlation with typicality are very close or even equal to 1 in the case of *art forms* ($\theta = .85$), *crimes* ($\theta = 1$), *emotions* ($\theta = .97$), and *media* ($\theta = 1$). The external features, resulting from the contrast categories of *media*, *emotions*, and *virtues* (twice), do not appear to influence the graded structure of the categories that Study 1 indicated to be well accounted for by the target features model. The θ values that yield optimal correlations with typicality do differ considerably from 1 for the categories of *diseases* ($\theta = .59$), *sciences* ($\theta = .62$), and *virtues* ($\theta = .09$). In the categories of *diseases* and *virtues*, this results in a considerable increase in the correlation with typicality from .58 ($p < .05$) to .74 ($p < .01$), and from .09 (*ns*) to .54 ($p < .05$), respectively. For *sciences*, the increase from .48 ($p < .05$) to .49 ($p < .05$) is rather small.

We employ AIC_c (Burnham & Anderson, 2002; Hurvich & Tsai, 1989) to compare the results of the original and the generalized polymorphous concept, instead of using statistical

---

[3] Other algorithms for representing similarity data, such as multidimensional scaling, could have been used to this end as well. The choice for an additive tree clustering procedure reflects considerations of visualization rather than of representation.

[4] We followed Lee (2001) in using the Bayesian information criterion (BIC, Schwarz, 1978) to indicate the number of internal nodes that provide the best compromise between the competing pressures of representational complexity and fit. By increasing the number of internal nodes to improve fit, one runs the risk of providing an account that is too complicated in light of the data's precision. The average of the standard deviations for each of the pooled cells in the final similarity matrix was taken as a measure of data precision.

Fig. 2 Correlation between graded structure and generalized polymorphous concept across θ values. Distinctive features originate from the most similar external category
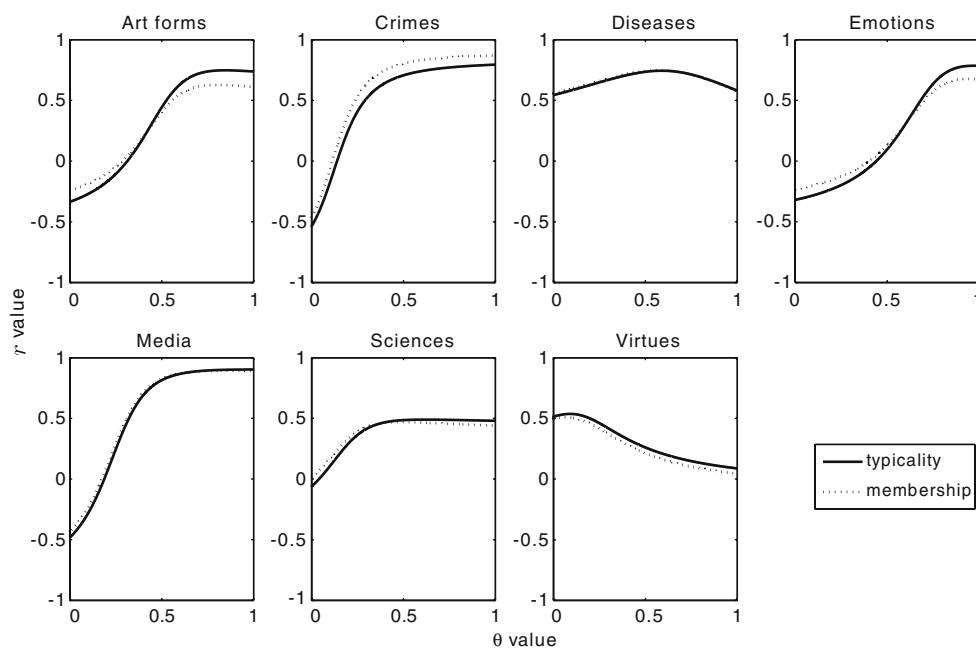
significance tests. AIC_c is a variant of Akaike's (1973) information criterion (AIC) with a small-sample bias-correction term, whereas the performance—in terms of Type I errors and power—of existing tests to compare dependent correlations is questionable in small samples such as the ones we have in the present study ($N = 15$; Hittner, May, & Silver, 2003; Wilcox & Tian, 2008). AIC_c has the added benefit over traditional significance tests in that it takes the additional parameter of the generalized model into consideration when comparing the results. In only two categories did AIC_c favor the model with external features over the target-features-only model. The respective AIC_c values were 10.24 and 12.36 for *diseases* and 26.37 and 27.53 for *virtues*. The improvement in correlation that the model with external features affords in *sciences* is not strong enough to favor it over the target-features-only model ($AIC_c = 43.99$ and $AIC_c = 40.36$, respectively).

Strikingly similar results are obtained when category membership constitutes the measure of gradedness. The optimal correlation between the generalized polymorphous concept and category membership is obtained for a θ value of .80 in the case of *art forms*, 1 in the case of *crimes*, .94 in the case of *emotions*, and .88 in the case of *media*. As was the case for typicality, these θ values are all very close or equal to 1, suggesting that it suffices to include target features to account for graded membership. Allowing the characteristic features of the contrast categories *emotions*, *art forms*, and *media* to contribute to the account of the graded membership of the exemplars of *diseases*, *sciences*, and *virtues* yields θ values that differ considerably from 1. $\theta = .58$ for *diseases*, $\theta = .49$ for *sciences*, and $\theta = .07$ for *virtues*. In the case of *diseases*, this change in θ improves the optimal correlation from .57 ($p < .05$) to .75 ($p < .01$).

In the case of *virtues*, the correlation with category membership increases from .04 (*ns*) to .51 ($p < .05$). As was the case for typicality, the increase in correlation with category membership for *sciences* from .44 ($p = .05$) to .47 ($p < .05$) is rather small. This suggests that the external features from the category of *art forms* do not influence the graded membership of the exemplars in *sciences*.

The AIC_c values support these observations. Only for the categories of *diseases* and *virtues* was AIC_c lower for the model with external feature information than for the model without: $AIC_c = -10.03$ vs. $AIC_c = -7.56$, and $AIC_c = -.88$ vs. $AIC_c = -.24$, respectively.

The pattern of results that was obtained with the generalized polymorphous concept model demonstrates a remarkable parallel to the pattern of results that was obtained with the original polymorphous concept model. The categories for which a target-feature-only approach fared well in Study 1 yielded an optimal θ value that was close or identical to 1 in Study 2. The categories for which the target-feature-only approach fared less well in Study 1 yielded an optimal θ value that differed considerably from 1. According to the model analyses, external features thus had a bigger part to play in the account of the graded structure of the latter categories. For two of them (*diseases* and *virtues*), the consideration of external features resulted in a more satisfying account of both typicality and category membership. Although the θ value that was obtained for the category of *sciences* also indicated a role for external features, this was not accompanied by a considerable increase in the correlation with typicality or category membership. We consider here two explanations for why this might be the case. The first pertains to the possibility that none of the abstract categories we are considering in

the present article constitutes an appropriate contrast for the category of *sciences*. The contrast selection procedure we employed is limited in that it merely indicates which category *of a given set* is most likely to constitute a contrast for a target category. The selected category will not necessarily act as a contrast. Current knowledge about the organization of abstract categories precludes a generative procedure that indicates likely suspects from among *all possible* contrasts. We will take up this point in the General Discussion, where we will suggest a number of manners in which one could start developing such a generative procedure. In the following section, we will concern ourselves with the possibility that characteristic features of more than one external category are required to warrant a considerable impact on the internal structure of the *sciences* category.

Study 3

There is no principled reason why only one external category should be expected to exert an influence on the target category. It is very likely that multiple contrast categories exert an influence on the graded structure of a target category. Evidence for this claim in the domain of concrete categories (reviewed in Verheyen et al., 2010) comes from the common procedure for eliciting contrasts (Malt & Johnson, 1992, see previous description), which generally yields more than one alternative to the suggested target category. Evidence for the existence of multiple contrasts in the domain of abstract categories comes from studies that have investigated the processing of abstract concepts through reference to others (i.e., metaphorical mapping). Many abstract concepts are found to be understood through several such references. Each of these references then provides part of the meaning of the abstract notion (Lakoff & Johnson, 1980). The *sciences* constitutes a prime example, with several well-documented metaphors (e.g., Banville, 1998; Christidou, Dimopoulos, & Koulaidis, 2004; Gerhart & Russell, 2004). Each of these highlights a particular aspect of the category's meaning. Other categories might also benefit from the inclusion of more than one contrast category. There exists, for instance, an entire literature devoted to the use of metaphors to document the manifestations of *diseases* (e.g., Koteyko, Brown, & Crawford, 2008; Skelton, Wearn, & Hobbs, 2002; Wallis & Nerlich, 2005). To investigate this possibility, we repeated the model analyses from Study 2, but instead of having the most similar alternative category act as a contrast, we had all six alternative categories take the role of contrasts.

Model analyses

We repeated the analyses using the generalized polymorphous concept model that we conducted in Study 2. The procedure was, in all respects, identical to that employed for Study 2, with the exception that external features could originate from any of the six alternatives to the target, not just from the category that was deemed most similar to the target in the additive tree clustering solution. To this end, the entire feature x exemplar matrix that was obtained in Study 1 was employed. It provided the exemplar feature vectors v_i that entered into Eq. (4). The category feature vector v_A in Eq. (4) held ones for those features that were generated in response to the target category label in the feature generation task, and zeros for those that were not. The bold lines in Fig. 3 display for each category the correlation of the generalized polymorphous concept with judged typicality across the entire range of θ. The dotted lines display the correlation with judged category membership. The results for $\theta = 1$ are necessarily the same as those that were obtained in Study 1 and Study 2. A θ value of 1 in the generalized polymorphous concept model indicates that only features that the target category shares with its exemplars are taken in consideration. This doesn't hinge on the choice of contrasting categories. Of interest is whether the optimal correlations that are obtained between the measures of gradedness and the generalized polymorphous concept model change as a function of this choice.

The results that are displayed in Fig. 3 closely resemble those of Fig. 2 in that the values of θ that yield the generalized polymorphous concept model's optimal account of typicality and category membership support the same division of categories we have seen in the previous studies. This conclusion should not come as a surprise in light of the manner in which contrast categories were chosen in Study 2. When the category that is found to be most similar to the target does not yield an effect, it is rather unlikely that categories that are less similar to the target (i.e., less suspect to provide external features through overlap of representations) will yield such an effect. According to the same rationale, it would be unlikely that categories that are less similar could completely undo the external feature effects found in Study 2, which originated from the most similar categories.

Although it appears, then, that for the majority of categories, the inclusion of multiple contrasting categories did not have a heavy impact on the account of category gradedness, two exceptions are of interest. In Study 2, it was established that the category gradedness of *virtues* was better accounted for by a measure of exemplar-category similarity that included both target and external features. Although the target-features-only model yielded correlations with typicality and category membership that were not significant, inclusion of external features originating from the category of *media* yielded correlations of .54 and .51 (both ps < .05). With the inclusion of multiple alternative categories, this increase became even more pronounced. The correlation with typicality changed from .09 (using target features only)

Fig. 3 Correlation between graded structure and generalized polymorphous concept across θ values. Distinctive features can originate from all external categories under consideration
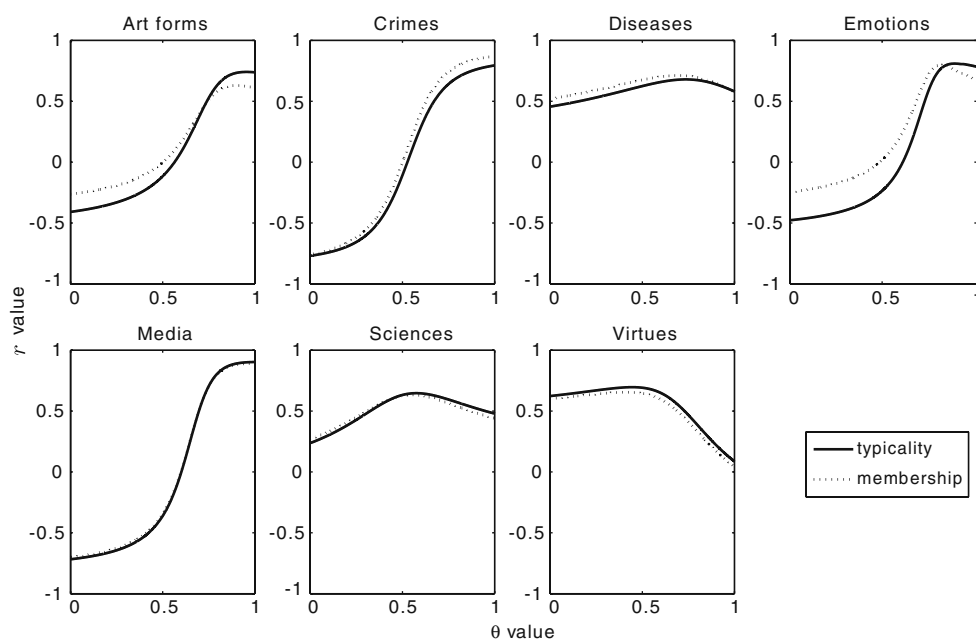
to .70 ($p < .01$). The correlation with category membership changed from .04 to .66 ($p < .01$). The corresponding AIC_c values favor the model with external features arising from all six contrast categories ($AIC_c = 21.80$ and $AIC_c = -4.80$) over the model with external features from *media* ($AIC_c = 26.37$ and $AIC_c = -.88$) and the model without external features ($AIC_c = 27.53$ and $AIC_c = z\text{-}.24$). In Study 2, the optimal θ for *sciences* was found to differ considerably from 1, but inclusion of external features originating from the category of *art forms* did not improve the correlation with typicality or category membership. The inclusion of multiple alternative categories did. The correlation with typicality changed from .48 ($p < .05$) to .65 ($p < .01$). The correlation with category membership changed from .44 ($p = .05$) to .63 ($p < .01$). The corresponding AIC_c values favor the model with external features arising from all six contrast categories ($AIC_c = 39.95$ and $AIC_c = 15.94$) over the model with external features from *art forms* ($AIC_c = 43.99$ and $AIC_c = 19.90$) and the model without external features ($AIC_c = 40.36$ and $AIC_c = 16.58$).

Other results, such as those for *diseases*, for instance, point toward the shortcomings of the procedure we have employed. In Study 2, the inclusion of external features originating from the category of *emotions* improved the correlation with typicality from .58 to .74 and with category membership from .57 to .75. In Study 3, the inclusion of external features originating from all six alternative categories increased these correlations with typicality and category membership to .68 and .71, respectively. The point is not that the increase after inclusion of multiple external categories should have been greater than after the inclusion of a single contrast, since we know there to be multiple aspects to the meaning of *diseases* (Koteyko et al.,

2008; Skelton et al., 2002; Wallis & Nerlich, 2005). After all, it is likely that none of the alternative categories we have considered (except for *emotions*) tap into these aspects. The point is that we observe a *decrease* in the correlation, which might be the result of error fitting. Similarly, for the categories of *art forms*, *emotions*, and *media*, we have observed θ values that are close to, but differ nonetheless, from 1. The corresponding increases in correlation observed for these categories are so small that we suspect them to reflect error fitting, rather than a small, but genuine, contribution of external features. The use of a model selection procedure such as AIC_c ensures that we will not attribute too much importance to θ values that are different from 1, but do not yield a substantial improvement of the correlations. Indeed, in these circumstances, AIC_c will favor the model with θ fixed at 1 over the model with a θ parameter that is free to vary, because the latter's improvement is too little in light of its increased flexibility. Ultimately however, to resolve discussions pertaining to what constitutes true evidence in favor of an external category's contribution to an abstract category's gradedness and what does not, we will need to move beyond the mere application of the generalized polymorphous concept model and include explanations of the occurrence or magnitude of external feature effects. The next study was an attempt to do so in that it identifies a correlate of the category differences we have been observing in the previous studies.

Study 4

The results of Hampton (1981) were replicated in Study 1: The extent to which feature commonality between exemplars

and category provides a satisfying account of graded category structure varied from one abstract category to the other. The correlation between category gradedness and the polymorphous concept, which measures the number of features that exemplars and category have in common, was less pronounced for the categories of *diseases*, *sciences*, and *virtues* than it was for the four other abstract categories that were included for study. The results of Study 2 and Study 3 suggested that abstract categories also vary with respect to their relational character. Incorporating features that target category exemplars share with external categories yielded an improved account of some abstract categories' graded structures, but not of all. The correlation between category gradedness and the generalized polymorphous concept, which takes both target and external feature information into account, was considerably higher than that between category gradedness and the original polymorphous concept for the categories of *diseases*, *sciences*, and *virtues*, but not for the four other ones. Small though our sample of categories might be, these results seem to suggest that the class of abstract categories is not a homogeneous one. To account for the graded structure of some of the abstract categories, it suffices to consider only those features that the exemplars and the category share in common. Other categories appear to be more relational in nature: To account for their graded structure, features that the exemplars share with one or more external categories need also be considered. This of course raises the question of what sets these categories apart. In the following section, we will evaluate the hypothesis that the categories differ with respect to abstractness.

Concreteness judgments

The procedure to obtain concreteness judgments was similar to that employed in the typicality and category membership judgment tasks. A group of 30 first-year psychology students of the University of Leuven received a booklet containing seven pages with, on each page, the 15 exemplars of one of the seven categories. The participants were asked to indicate for every item on the page how abstract they found it to be. They were required to provide their answers by indicating a value on a scale ranging from 1 (*referring to highly abstract items, that cannot be perceived with our senses*) to 7 (*referring to very concrete items, that are easy to see, hear, or feel*). These instructions were taken from Gernsbacher (1984). If participants did not know one of the presented exemplars, they were asked to indicate this by drawing a circle around it. Two different presentation orders for the categories and three different presentation orders for the exemplars in a category were employed. Figure 4 holds boxplots of the concreteness judgments for the exemplars of the seven abstract categories. The middle band of the boxes represents the average concreteness of the category exemplars.
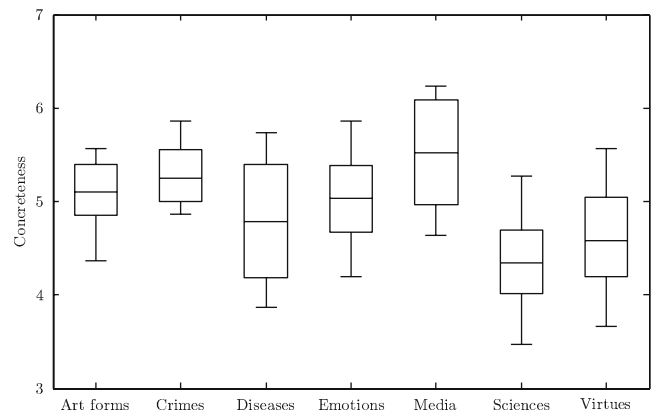

Fig. 4 Boxplots of the concreteness judgments for the exemplars of the seven abstract categories

The concreteness judgments for every exemplar were averaged across participants and were subjected to an ANOVA with category as a between-exemplars variable. A significant effect of category was found, $F(6, 98) = 8.75$, $MS_e = 2.43$, $p < .0001$, indicating that participants did not find the seven categories equally abstract. A planned comparison showed that the four abstract categories that did not take external features to account for graded category structure (*art forms*, *crimes*, *emotions*, and *media*) were judged to be less abstract than those that did (*diseases*, *sciences*, and *virtues*), $F(1, 98) = 39.56$, $MS_e = 10.96$, $p < .0001$. Tukey post hoc comparisons of the seven categories revealed that neither the four "target only" categories nor the three "target and external" categories differed from one another in terms of judged concreteness. The category of *sciences* ($M = 4.34$, $SD = .47$) was, however, found to be more abstract than the categories of *art forms* ($M = 5.10$, $SD = .47$), *crimes* ($M = 5.25$, $SD = .45$), *emotions* ($M = 5.03$, $SD = .48$), and *media* ($M = 5.52$, $SD = .59$). The exemplars in the category of *virtues* ($M = 4.58$, $SD = .55$) were judged significantly more abstract than the exemplars in *crimes* and *media*. The exemplars in the category of *diseases* ($M = 4.78$, $SD = .64$) were judged significantly more abstract than those in the category of *media*.

This study is not the first one to demonstrate that judgments of concreteness reveal that abstract concepts do not form a homogeneous class (see, e.g., Altarriba, Bauer, & Benvenuto, 1999; Wiemer-Hastings, Krug, & Xu, 2001). The results are, however, novel in that they show that the abstract categories that are judged to be most abstract are the ones that benefit from the inclusion of external features in addition to target features to account for their graded structure. It appears, then, that the abstract categories we found to be most relational in nature are also the ones that are most abstract. Such a result makes sense in that the processing of highly abstract categories is likely to benefit from references to other, less abstract categories (Lakoff & Johnson, 1980).

General discussion

Although there is evidence to suggest that abstract and concrete categories are processed differently (Lakoff & Johnson, 1980; Lakoff & Turner, 1989) and are associated with different types of features (Barsalou & Wiemer-Hastings, 2005; Wiemer-Hastings & Xu, 2005), the results of the present studies provide further evidence that, with respect to gradedness, abstract categories and concrete categories are alike. Both types of categories exhibit graded structure in the form of exemplar generation frequencies, typicality judgments, and category membership judgments. Additionally, the model analyses in the present study (along with the analyses in Dry & Storms, 2010) suggest that graded structure in these categories can be explained in the same manner. Specifically, it appears that for both category types, typicality and membership gradations are a reflection of the underlying similarity structure of the category, with exemplars that are highly similar to the category representation being more representative than exemplars with a low degree of similarity to the category representation.

The latter conclusion builds on the work of Hampton (1981), which established that the extent to which feature commonality between exemplars and category provides a satisfying account of a graded category structure varies from one abstract category to the other. Following the suggestion that abstract categories are relational in nature (Goldstone, 1996; Wiemer-Hastings & Xu, 2005), we argued that extending exemplar-category similarity to include features that exemplars share with external categories might yield an improved account of abstract categories' graded structures. The generalized polymorphous concept model, which incorporates both types of features, was found to improve the account of typicality and category membership in the three categories that didn't fare well under the feature commonality account. These three categories thus fulfill the requirements for being called "relational categories" (Gentner, 1981; Markman & Stilwell, 2001). They combine a relatively weak internal structure with strong links to external categories. Our results resemble those of Dry and Storms (2010), who applied the generalized polymorphous concept model to account for concrete categories' graded structures. They too found that the contribution of external feature information varied from one category to the other.

Rather than supporting a broad distinction between concrete and abstract categories, the present results, combined with those of Dry and Storms (2010), suggest that these categories can be organized along a continuum. Categories with a graded structure that is not influenced by external categories would be found on one end of this continuum. Categories with a graded structure that is strongly influenced by external categories would be found on the other end. We do not argue for a new distinction (between nonrelational and relational categories), but believe that the results support the idea of an isolatedness–interrelatedness continuum of the kind that was proposed by Goldstone (1996). Indeed, both in our studies on abstract categories and in those of Dry and Storms on concrete categories, there appears to exist considerable variety among categories in the extent to which their organization is influenced by external features. Similarity to the category representation, expressed as a weighted combination of target and external features, might be at the origin of graded structure in both types of categories, but intercategory differences herein remain. Much like the question of what affects the relative contribution of common and distinctive features in similarity ratings (Gati & Tversky, 1984; Tversky, 1977), what might account for these differences found amid concrete and abstract categories is a question well deserving further study.

The results of Study 4 argue that the abstractness of the constituting exemplars might be involved in categories' organization along the isolatedness–interrelatedness continuum. The categories that were least influenced by external categories were found to be the least abstract. Crutch and Warrington (2005) have already argued that concreteness–abstractness constitutes a continuum ranging between highly concrete concepts and highly abstract concepts. It would be interesting to see how well categories' organization along this continuum accords with that along the isolated–interrelated continuum. Another straightforward step to take would be to take up the processing and structural differences that have been found by treating abstract and concrete categories as two distinct classes and to investigate how much variability *within* each class of categories exists with respect to these differences. The generalized polymorphous concept model could play a supporting role in such investigations. Differences in supposed explanatory variables for interrelatedness should be reflected in the results of the generalized polymorphous concept model. The mean concreteness differences that were obtained in Study 4, for instance, support the same distinction of categories that we made on the basis of the model analyses in Studies 2 and 3.

Of course, to allow such investigations across categories that vary considerably along the concreteness–abstractness continuum, we are in need of methods that allow us to elicit potential contrast categories for both highly concrete and highly abstract categories. In the introduction to Study 2, we had already hinted that the procedures that are commonly employed for concrete categories cannot readily be applied to abstract categories. The main reason for this is that many concrete categories are hierarchically organized, whereas many abstract categories are not (Crutch & Warrington, 2005; Hampton, 1981). When combined with analyses of large text corpora, exploratory studies of the kind reported in the present article and in that of Dry and Storms

(2010) might bring us one step closer to a procedure to elicit contrast categories that is applicable to categories across the concreteness–abstractness continuum. First, the generalized polymorphous concept model could be employed to reveal categories that are related. Subsequent analyses of the occurrences of these categories in text might reveal context elements that are predictive of the interrelatedness of the categories. These can then be used as a heuristic for identifying other related categories. Of course, these context elements need not be the same across the concreteness–abstractness continuum. The work by Wiemer-Hastings (Wiemer-Hastings, 1998; Wiemer-Hastings & Graesser, 2000) on procedures to determine the relevance of various context elements for the meaning of abstract categories suggests that at the abstract end of the continuum, the types of verbs with which the categories co-occur might prove to be of particular importance.

Familiarity and multiple senses

We have shown that the generalized polymorphous concept model provides a satisfying account of the graded structure of abstract categories by utilizing both target and external feature information. However, as with the well-defined categories in Larochelle et al. (2000), it is possible that the present results are due to a confounding of representativeness and familiarity. In particular, Hampton (1981) speculated that the familiarity of exemplars might influence the graded structure of abstract categories to a larger extent than it does that of concrete categories. (For demonstrations of the effect that familiarity can have on typicality judgments in concrete categories, see Malt & Smith, 1982, and Hampton & Gardiner, 1983.) Specifically, it was argued that if some abstract concepts have no single definition, then the graded structure present in these categories might reflect the frequency with which the category members are encountered (e.g., familiarity) rather than the similarity relations between the exemplars and the category representation(s). In accordance with the view that abstract concepts might have more distinct meanings than concrete concepts is the finding by Galbraith and Underwood (1973) that abstract words are judged to appear in a greater variety of contexts than are concrete words. When the variety of contexts is derived from a text corpus, instead of being judged, the same difference is obtained (Hoffman, Rogers, & Lambon Ralph, 2011). Accordingly, when sentence or paragraph contexts are provided in semantic tasks, response time differences between abstract and concrete stimuli disappear (Schwanenflugel, Harnishfeger, & Stowe, 1988; Schwanenflugel & Shoben, 1983; Schwanenflugel & Stowe, 1989). The work on metaphorical mapping in the processing of abstract categories (Lakoff & Johnson, 1980) also argues that there might be distinct parts to the meaning of abstract

concepts. The results of Study 3, in which multiple external categories were allowed to influence the account of the abstract categories' graded structures, at least support the idea that there are distinct parts to the meaning of some of the categories we have been considering.

To test whether the frequency with which raters come into contact with, see, or interact with an exemplar does not present a potential confound or source of noise, we gathered familiarity judgments for each of the exemplars included in the present article, used regression analysis to partial out the shared variance between the gradedness and familiarity judgments, and then subjected the resulting residuals to the model analyses employed in Study 3 (for a description of the procedure used to obtain the familiarity judgments, see Stukken, Verheyen, Dry, & Storms, 2009). If the graded structure in the abstract categories is indeed a reflection of exemplar familiarity, then the generalized polymorphous concept model should fare less well. However, the procedure did not have a heavy impact on the results. If anything, the evidence for an influence of external features on abstract categories' graded structures became more apparent. These results are most likely due to the weak relationship between the familiarity and gradedness judgments for the majority of the categories. Only for *art forms* ($r = .53$, typicality) and *media* ($r = .60$, typicality; and $r = .61$, category membership) did the correlation reach significance at the .05 level (two-tailed t). It would appear that familiarity does not have a strong impact on gradedness and can therefore not be held responsible for the influence of external feature information on abstract categories' graded structures. This result adds further weight to the suggestion that the graded structure in abstract categories is based on exemplar-category similarity.

Features and feature weights

It is doubtful that the established contribution of external features hinges on the nature of the features that were employed to implement the generalized polymorphous concept model. Following Hampton (1981), we used features that were generated toward the category label instead of toward the exemplars themselves. Although exemplar features provide for a richer representation than category features do, this does not have a heavy impact on accounts of gradedness. Dry and Storms (2010) demonstrated how, in concrete categories, the generalized polymorphous concept account of typicality using category-level features is not inferior to that of the generalized polymorphous concept with exemplar level features. Nor does it seem to matter whether or not features are given equal weight. Using the individual feature x exemplar matrices as input to the model analyses and averaging across the resulting correlations yielded results that were qualitatively similar to the ones displayed in Figs. 2 and 3. Similar results were also obtained

when the five feature x exemplar matrices were combined into a single matrix by determining the majority judgment for each cell and this binary matrix was input to the model analyses. Whether or not features are weighted according to the number of participants that endorsed them thus does not seem to have a heavy impact on our results.

Conclusion

The results of the present study suggest that the relationship between graded structure and exemplar-category similarity in abstract categories is more pervasive than previously thought. A weighted combination of target features (features that exemplars share in common with the target category) and external features (features that exemplars share in common with one or more external categories) was found to reflect the representativeness of the abstract categories' exemplars. This result was obtained both when typicality and category membership were adopted as measures of gradedness. The contribution of target features was more pronounced among the more concrete of the studied categories. External features were weighted more heavily among the more abstract categories under investigation. It would appear, then, that concreteness–abstractness is one determinant of categories' organization along a continuum extending from isolated to interrelated categories.

Appendix Selected exemplars per category, in descending order of generation frequency

Art Forms	Crimes	Diseases	Emotions	Media	Sciences	Virtues
painting	murder	influenza	joy	television	chemistry	friendship
sculpting	rape	Parkinson's disease	grief	radio	biology	honesty
music	fraud	cold	fear	newspapers	physics	helpfulness
dancing	hold-up	Alzheimer's	happiness	Internet	mathematics	fidelity
architecture	burglary	laryngitis	rage	advertising	geography	deference
drawing	carjacking	malaria	love	telephone service	astronomy	kindness
photography	vandalism	measles	disappointment	computers	sociology	modesty
graffiti	abduction	asthma	surprise	posters	educational science	empathy
poetry	battery	syphilis	shyness	commentaries	anatomy	prudence
fashion	stalking	rheumatism	disillusionment	teletext	genetics	ambition
molding	arson	anorexia nervosa	nervousness	compact discs	economics	courage
musical	paedophilia	migraine	loneliness	photography	information science	courtesy
glassblowing	incest	diabetes	pride	faxing	history	beauty
flower arrangement	illegal dumping	dementia	boredom	blogging	linguistics	self-development
comics	driving against traffic	depression	lethargy	telegrams	dentistry	openness

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, and Computers, 31*, 578–602.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*, 263–308.

Banville, J. (1998). Beauty, charm, and strangeness: science as metaphor. *Science, 281*, 40–41.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211–227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 629–654.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New York, NY: Cambridge University Press.

Bouma, G., van Noord, G., & Malouf, R. (2001). Alpino: wide-coverage computational analysis of Dutch. *Language and Computers, 37*, 45–59.

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11*, 617–660.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer.

Christidou, V., Dimopoulos, K., & Koulaidis, V. (2004). Constructing social representations of science and technology: the role of metaphors in the press and popular scientific magazines. *Public Understanding of Science, 13*, 347–362.

Church, K. W., & Hanks, P. (1991). Word association norms, mutual information and lexicography. *Computational Linguistics, 16*, 22–29.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain, 128*, 615–627.

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

Dry, M. J., & Storms, G. (2010). Features of graded category structure: generalizing the family resemblance and polymorphous concept models. *Acta Psychologica, 133*, 244–255.

Estes, Z. (2004). Confidence and gradedness in semantic categorization: definitely somewhat artifactual, maybe absolutely natural. *Psychonomic Bulletin & Review, 11*, 1041–1047.

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition, 1*, 56–60.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgements. *Cognitive Psychology, 16*, 341–370.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4*, 161–178.

Gerhart, M., & Russell, A. M. (2004). Metaphor and thinking in science and religion. *Zygon, 39*, 13–38.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and ortography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256–281.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24*, 608–628.

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

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition, 9*, 149–156.

Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: a correlational analysis of normative data. *British Journal of Psychology, 74*, 491–516.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology, 130*, 149–168.

Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience.*

Hough, M. S., & Pierce, R. S. (1989). Exemplar verification for common and ad hoc categories in aphasia. In *Clinical aphasiology* (Vol. 19, pp. 139–150). Austin, TX: Pro-Ed.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*, 297–307.

Koteyko, N., Brown, B., & Crawford, P. (2008). The dead parrot and the dying swan: The role of metaphor scenarios in the UK press coverage of avian flu in 2005–2006. *Metaphor and Symbol, 23*, 242–261.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*, 195–208.

Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor.* Chicago, IL: University of Chicago Press.

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.

Larochelle, S., Richard, S., & Soulières, I. (2000). What some effects might not be: the time to verify membership in "well-defined" categories. *Quarterly Journal of Experimental Psychology, 53A*, 929–961.

Lee, M. D. (1999). *Algorithms for representing similarity data* (Defence Science and Technology Organisation Research Report, DSTO-RR-0152).

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling models 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.

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

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10*, 69–75.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence, 13*, 329–358.

Martin, J. D., & Billman, D. O. (1994). Acquiring and combining overlapping concepts. *Machine Learning, 16*, 121–155.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121–147.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society, 7*, 283–284.

Navarro, D. J., & Lee, M. D. (2004). Common and distinctive features in stimulus similarity: a modified version of the contrast model. *Psychonomic Bulletin & Review, 11*, 961–974.

Osherson, D., & Smith, E. E. (1997). On typicality and vagueness. *Cognition, 64*, 189–206.

Pado, S., & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics, 33*, 161–199.

Pinker, S. (1997). *How the mind works.* New York, NY: Penguin.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniardou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). Cambridge, England: Cambridge University Press.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York, NY: Academic.

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

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319–345.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*, 499–520.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 82–102.

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly, 24*, 114–126.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210*, 390–398.

Skelton, J. R., Wearn, A. M., & Hobbs, F. D. R. (2002). A concordance-based study of metaphoric expressions used by general practitioners and patients in consultation. *British Journal of General Practice, 52*, 114–118.

Stukken, L., Verheyen, S., Dry, M. J., & Storms, G. (2009). A new investigation of the nature of abstract categories. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2438–2443). Austin, TX: Cognitive Science Society.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

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

Verheyen, S., De Deyne, S., Dry, M. J., & Storms, G. (2010). *Uncovering contrast categories in categorization with a probabilistic threshold model*. Manuscript submitted for publication.

Wallis, P., & Nerlich, B. (2005). Disease metaphors in new epidemics: the UK media framing of the 2003 SARS epidemic. *Social Science & Medicine, 60*, 2629–2639.

Wiemer-Hastings, K. (1998). Abstract noun classification: using a neural network to match word context and word meaning. *Behavior Research Methods, Instruments, & Computers, 30*, 264–271.

Wiemer-Hastings, K., & Graesser, A. C. (2000). Conceptually representing abstract concepts with abstract structures. In L. Gleitman & A. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 983–988). Hillsdale, NJ: Erlbaum.

Wiemer-Hastings, K., Krug, J. D., & Xu, X. (2001). Imagery, context availability, contextual constraint and abstractedness. In J. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 1106–1111). Mahwah, NJ: Erlbaum.

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science, 29*, 719–736.

Wilcox, R. R., & Tian, T. (2008). Comparing dependent correlations. *Journal of General Psychology, 135*, 105–112.