Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 6, 1515–1531

0278-7393/11/$12.00 DOI: 10.1037/a0024431

Uncovering Contrast Categories in Categorization With a Probabilistic Threshold Model

Steven Verheyen and Simon De Deyne
University of Leuven

Matthew J. Dry
University of Adelaide

Gert Storms
University of Leuven

A contrast category effect on categorization occurs when the decision to apply a category term to an entity not only involves a comparison between the entity and the target category but is also influenced by a comparison of the entity with 1 or more alternative categories from the same domain as the target. Establishing a contrast category effect on categorization in natural language categories has proven to be laborious, especially when the categories concerned are natural kinds situated at the superordinate level of abstraction. We conducted 3 studies with these categories to look for an influence on categorization of both similarity to the target category and similarity to a contrast category. The results are analyzed with a probabilistic threshold model that assumes categorization decisions arise from the placement of threshold criteria by individual categorizers along a single scale that holds the experimental stimuli. The stimuli's positions along the scale are shown to be influenced by similarity to both target and contrast. These findings suggest that the prevalence of contrast category effects on categorization might have been underestimated. Additional analyses demonstrate how the proposed model can be employed in future studies to systematically investigate the origins of contrast category effects on categorization.

Keywords: natural language categories, contrast categories, categorization, typicality, similarity

This article is concerned with a common act we all undertake numerous times a day to facilitate understanding and communication: deciding the category membership of a specific entity with which we are confronted. We ask whether this decision is governed by information pertaining to the entity (e.g., *avocado*) and the target category (***fruits***) solely, or whether information pertaining to possible alternative categories (***vegetables***) influences the decision as well. To answer this question we introduce a formal model that is tailored to the inter-individual categorization differences that are characteristic of natural language categories. The model allows for a test of the independent contributions of items' similarity to the target category and items' similarity to a contrast category. The strategy we employ is to apply the model to categorization data for those categories that, based on the existing literature, are least likely to be influenced by contrasting categories. If contrast category effects can be shown to exist among these categories, they might be more pervasive than has previously been suggested. Such a result would also imply that the interrelatedness of natural language categories' representations has been underestimated. We start by providing an overview of the existing literature on contrast category effects in categorization.

Demonstrations of Contrast Category Effects on Categorization

A contrast category effect on categorization occurs when the decision to apply a category term to an entity not only involves a comparison between the entity and the target category but is also influenced by a comparison of the entity with one or more alternative categories from the same domain as the target. While categories are thought to vary along a continuum of interrelatedness (Goldstone, 1996), contrast category effects have proven elusive in categorization tasks involving natural language categories. Following an elaborate set of studies, Verbeemen, Vanoverberghe, Storms, and Ruts (2001) concluded that not many natural language categories reside at the interrelated pole of the continuum. Verbeemen et al. set out to verify whether category exemplars' speed of categorization was influenced by their similarity to both the target category and one or more contrast categories. In accordance with the family resemblance notion by Rosch and

This article was published Online First July 18, 2011.

Steven Verheyen, Simon De Deyne, and Gert Storms, Department of Psychology, University of Leuven, Leuven, Belgium; Matthew J. Dry, University of Adelaide, Adelaide, South Australia, Australia.

Steven Verheyen is a research assistant at the Research Foundation—Flanders. Simon De Deyne is a postdoctoral fellow at the Research Foundation—Flanders. The presented data can be found at Steven Verheyen's website at http://ppw.kuleuven.be/concat/

We would like to thank Eef Ameel, Bieke Bollen, James A. Hampton, Daniel Heussen, Wolf Vanpaemel, and Wouter Voorspoels for their valuable input. We thank Michael D. Lee and Francis Tuerlinckx for sharing their knowledge on Bayesian model estimation and Loes Stukken for her assistance in gathering the required data.

Correspondence concerning this article should be addressed to Steven Verheyen, Department of Psychology, University of Leuven, Tiensestraat 102, Box 3721, BE-3000 Leuven. E-mail: steven.verheyen@psy.kuleuven.be

Mervis (1975), the more an exemplar was perceived as similar to the target category *and* dissimilar to a contrast category, the shorter its classification response time was expected to be. Rosch and Mervis never actually confirmed whether the extents to which an item demonstrates feature overlap with the exemplars of the target category and with the exemplars of a contrast category contribute independently to the item's perceived exemplariness. Verbeemen et al. attempted such a confirmation by regressing item classification response time on both item variables. The studies they conducted yielded eight opportunities to demonstrate independent contributions by the two variables. None of these yielded a significant contribution of the items' similarity to a contrast category over and above that of the items' similarity to the target category. When they used the items' rated similarity toward the 10 most frequently generated exemplars of target and contrast category as alternative operationalizations of the item variables, one of the eight conducted regression analyses yielded a significant contribution of both variables.

Hampton, Estes, and Simmons (2005) conducted a categorization task that involved the natural language categories ***cats*** and ***dogs***. They found participants were more likely to endorse an atypical stimulus as a category member when it followed a stimulus from the contrasting category than when it followed a stimulus from the target category. This finding suggests that dissimilarity to members of contrast categories contributes over and above similarity to target categories' members in categorization, but the result might be somewhat different in nature than that sought after by Verbeemen et al. (2001). While the finding by Hampton et al. pertains mainly to the immediate context in which categorization decisions are made (i.e., the items and corresponding categorization decisions that precede the decision for the current item), the question that Verbeemen et al. asked pertains to the influence of alternative categories on categorization decisions, regardless of the context in which they are made. That is, they ask to what degree category representations are interrelated.

This rather limited empirical evidence for contrast category effects on categorization is at odds with the multitude of studies that have reported contrast category effects on other category-related decisions, like judgments about items' representativeness (e.g., Ameel & Storms, 2006; Dry & Storms, 2010; Verheyen, Stukken, De Deyne, Dry, & Storms, 2011; Voorspoels, Vanpaemel, & Storms, 2008, in press). This is peculiar, as one would expect that the contribution of contrast categories is more pronounced for judgments of category membership than for judgments of representativeness. Indeed, while the question of category membership implies that the item might belong to other categories than the target, category membership is more or less assumed in the question of how representative an item is of a target category.

Moreover, the apparent absence of contrast category effects on categorization in natural language categories is at odds with the importance that is attributed to the effect in the artificial category learning literature. This literature includes empirical demonstrations of how the similarity of an item to one category can affect its perceived membership in another category. Palmeri and Nosofsky (2001), for instance, had participants in three perceptual categorization tasks learn to discern artificially created categories. Palmeri and Nosofsky demonstrated that the items that were categorized most accurately were the ones that resided at the target category's boundary most remote from the other categories involved in the task. The items that were considered clear category members thus were high in similarity to the target category's instances *and* low in similarity to the contrasting categories' instances (see also Goldstone, Steyvers, & Rogosky, 2003). The artificial category learning literature also includes various models that rely on Luce's (1959) choice rule to convert items' similarities to categories into response probabilities (e.g., the generalized context model, Nosofsky, 1984; ALCOVE, Kruschke, 1992). According to these models, the probability of a positive categorization decision increases with items' similarity to the target category and decreases with items' similarity to one or more alternative categories. Surely then, if these findings have any bearing on categorization involving natural language categories, the membership of at least some natural language categories should be affected by contrast categories.

Rather than turning toward the natural language categories that are most likely to be affected by contrast categories, however, the following section identifies those that are least likely to demonstrate contrast category effects. If contrast category effects on categorization can be shown to exist among the latter categories, the prevalence of the phenomenon must almost certainly have been underestimated.

Unlikely Candidates for Contrast Category Effects

Which categories are least likely to reside at the interrelated pole of the isolated–interrelated continuum? Markman and Wisniewski (1997) have argued that contrast category effects are harder to show at the superordinate level of natural language categories than at the basic level. Unlike basic-level categories, superordinate categories would share little or no representational structure. This would hamper the retrieval of other categories at the superordinate level that might function as contrasts. In line with this argument, Markman and Wisniewski have reported difficulties in eliciting contrasts for superordinate categories, but not for basic-level categories. Following the example of Malt and Johnson (1992), Markman and Wisniewski 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*** was incorrect. At that point they were asked for a second guess at the entity's identity. If a consistent answer is provided as a second guess, it is considered to be a contrast category of category ***X***. No consistent responses emerged for the superordinate categories, however.

More recently, Verbeemen et al. (2001) showed that at least for some superordinate categories a contrast can be generated reliably. However, for only one of these categories did they manage to demonstrate a reliable effect on speed of classification of the items' similarity to the generated contrast over and above that of their similarity to the target category. Clearly, it is not because a category can be reliably generated in response to another that it will prove to be the former's contrast in categorization. Hampton et al. (2005) did obtain results that were reminiscent of a contrast category effect on categorization.[1] They made use of the animal categories ***cat*** and ***dog***, which Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) established reside at the basic level of abstraction. Taken together, these results indeed support the impression that contrast category effects are more likely to emerge at the basic level of abstraction than at the superordinate level.

According to the same rationale, categories composed of natural items should be less prone to demonstrate contrast category effects than categories of artifacts (e.g., Goldstone, 1996). There is ample evidence to suggest that natural categories show less representational overlap than artifact categories do. Perhaps the most compelling demonstration of this was provided by Ceulemans and Storms (2010), who studied the latent structure of exemplar-by-feature applicability matrices. They compiled two applicability matrices, one for the natural domain and one for the artifact domain. The rows of each matrix were made up of a large selection of the corresponding domain's exemplars, while those features participants had deemed characteristic of these exemplars constituted the columns of the matrix. The matrix entries indicated whether or not the features applied to the exemplars. Ceulemans and Storms found the applicability matrix of natural items to be organized into mutually exclusive classes. They found the latent classes constituting the artifact applicability matrix to overlap.

Other work corroborates the idea that representational overlap is typical of artifact categories but not of natural categories. When asked to judge whether an item described as "halfway between" two categories (a) was probably one or the other, (b) could be called either one, or (c) couldn't be either, participants in a study by Malt (1990) generally opted for the first alternative for natural items, while preferring the second alternative for many artifacts. In addition, in exemplar generation tasks it is not uncommon for artifacts to be produced in response to several target categories, while natural items, on the other hand, tend to be generated toward a single target category only (De Deyne et al., 2008). Studies on free naming and category verification also indicate that artifact items afford multiple category labels (Malt & Sloman, 2004; Ruts, Storms, & Hampton, 2004). Moreover, a post hoc analysis of category verification data by Hampton (1998) suggested that underestimates in the prediction of categorization probability from items' typicality were often due to the items' membership in another category in the case of artifacts, but not in the case of natural items.

It is noteworthy that Verbeemen et al. (2001) included twice as many natural categories as artifact categories. Nevertheless, the one analysis they reported that yielded a contrast category effect involved artifact categories. The category of ***tools*** proved to act as a contrast category for the category of ***kitchen utensils***.

Analyzing Categorization Data With a Probabilistic Threshold Model

Any account of categorization that is concerned with natural language categories has to address the observation that people disagree about the items they feel should be endorsed as category members (McCloskey & Glucksberg, 1978). The threshold theory (Hampton, 1998, 2007) appears to meet this requirement. It assumes a latent dimension—common to all categorizers—along which the items are organized. Individual categorizers can allocate their categorization criteria at different locations along this latent dimension. Each of these criteria or *thresholds* divides the dimension: Items that are positioned on one side of the threshold are not considered members of the target category; the items that are positioned on the other side of the threshold are. Categorizers with different categorization thresholds differ with respect to the items they consider category members; they divide the dimension that holds the items in a different manner.

Verheyen, Hampton, and Storms (2010) have provided a formalization of the threshold theory by noting its commonalities with a model from the item response literature. They have shown how the aim of the Rasch model (Rasch, 1960; Thissen & Steinberg, 1986) to estimate participants' varying proficiency with regard to questions of varying difficulty translates easily to the threshold theory's goal to elucidate individuals' varying criteria in endorsing various items as category members. The model provides a good account of the inter-individual differences that are found in categorization with respect to natural language categories of a diverse nature, including natural kind categories, artifact categories, and abstract categories (Verheyen et al., 2010).

The Probabilistic Threshold Model

In line with the threshold theory's account of categorization, the Rasch model employs the structure contained in binary person-by-item categorization data to award both persons and items a position along a single scale. (Note that binary categorization decisions Y can arise from an ***A*** vs. ***not A*** task or from an ***A*** vs. ***B*** task. In the former case we refer to ***A*** as the target category. In the latter case one of the categories can arbitrarily be denoted the target category.) An item i's position along the scale is indicated by β_i. The more an item is endorsed as a target category member, the higher on the scale the model positions it (i.e., a higher β_i value). A person p's position along the scale is indicated by θ_p. It represents the degree of liberalness/conservatism the person displays when making categorization decisions. Low values of θ_p indicate rather liberal categorizers who consider many of the presented items to be target category members. High values of θ_p characterize more conservative categorizers who include fewer items in the target category. The relative position of β_i and θ_p along a logit scale determines the probability that person p will endorse item i as a member of the target category:

$$\Pr(Y_{pi} = 1) = \frac{e^{(\beta_i - \theta_p)}}{1 + e^{(\beta_i - \theta_p)}}. \quad (1)$$

Equation 1 expresses that the more the item's position exceeds the categorizer's position along the scale, the more likely it becomes that the item will be endorsed, and vice versa. It is therefore appropriate to interpret the θ_ps as categorization thresholds. To see this, consider the example in Figure 1. It displays the categorization threshold θ_p of a particular person and the positions β_i and β_j of two items along the latent scale (horizontal axis). The black curves in the figure indicate how the probability of endorsing the items as target category members changes as a function of θ_p.

[1] We use the term *reminiscent* because—as we already stated in the introduction—the effect might be of a different nature than that observed by Verbeemen et al. (2001), which is the one we are pursuing.

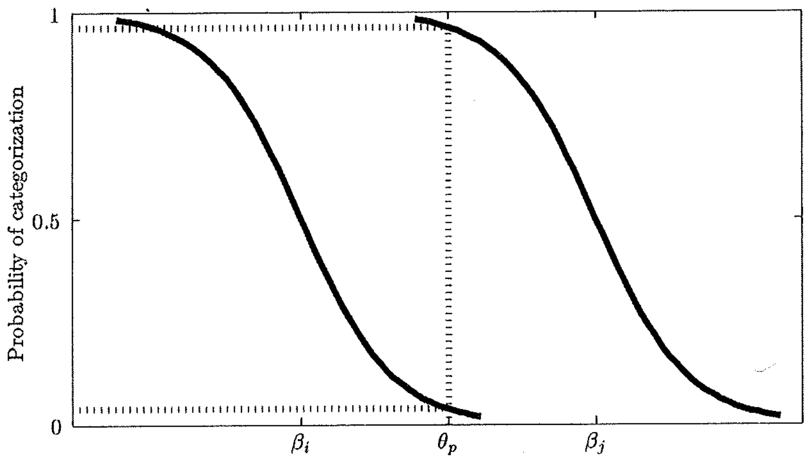

Figure 1. Illustration of the probabilistic threshold model.

In Figure 1 person p's threshold is located in between the two items. As β_j surpasses θ_p, item j has a high probability of being endorsed by p. The dashed vertical line at position θ_p in Figure 1 crosses the black response curve associated with β_j close to a categorization probability of 1. The probability of categorization is clearly much smaller for item i. β_i does not surpass θ_p and therefore has a low categorization probability associated with it. The dashed vertical line at position θ_p crosses the black response curve associated with β_i close to a categorization probability of 0.

Now imagine a person whose threshold is located to the left of both items. Both the items' βs then surpass θ_p, and both would have high categorization probabilities associated with them. With respect to these two items, this person can be considered a far more liberal categorizer than the person who is positioned such that neither β_i nor β_j surpasses the corresponding θ. In the latter case, both i and j would have a low probability of being endorsed.

The Rasch model can be considered a probabilistic version of the threshold theory. In the model θ_p acts like a threshold criterion, but a less rigorous one than the one proposed in the original threshold theory, where it made a clear-cut distinction between items that belong to the target category and items that do not. In the Rasch model a categorization decision is considered the outcome of a Bernoulli trial. Depending on the probabilities the Rasch model associates with both outcomes, the categorization outcome for a particular person–item combination needn't always be the same. When the person is asked to categorize the item again on a different occasion, her answer may differ from the one that was first produced. This property of the model allows it to account for both intra- and inter-individual differences in categorization in a single explanatory framework (see Verheyen et al., 2010, for details).

Its representational assumption sets the model apart from other accounts of categorization. Many existing categorization models assume that items are represented in a multidimensional space (see Ashby & Maddox, 1993, for an overview). The categorization processes they propose operate upon the information contained in these multidimensional representations to account for individuals' categorization decisions. A Rasch model application, on the other hand, doesn't require information beyond the categorization data to be fit. The model extracts its representation from the available categorization decisions. It uses the information contained in the binary person-by-item categorization data to organize the items along a single scale (i.e., to establish the β_i values). This organization provides for the best account of the categorization data under the Rasch model assumptions. Of course, what one is left with then is the interpretation of the β_is. According to the threshold theory, the scale represents a similarity metric with the β_is reflecting the similarity of the items' representation to the target category's representation (Hampton, 1998, 2007). The further along the scale an item is positioned, the higher its similarity to the target category is assumed to be. In Figure 1, for instance, the fact that item i is located lower on the scale than item j is thought to indicate that i is less similar to the target category than j. An individual's categorization decision is then assumed to involve the determination of the amount of similarity required to decide in favor of category membership (i.e., the placement of a threshold criterion). The low threshold value θ_p of liberal categorizers can then be taken to indicate that for them a modest degree of similarity suffices to conclude category membership. The high θ_p values that characterize conservative categorizers are then assumed to indicate that these categorizers require extensive similarity between item and category to conclude category membership. Clearly, the evaluation of such assumptions *does* require information beyond that contained in the categorization data. To interpret the nature of the scale that underlies people's categorization decisions one may want to bring external information into the model. In order to achieve this it is necessary to employ an extension of the model.

Extending the Probabilistic Threshold Model

The threshold theory has received considerable support by the successful application of the Rasch model to categorization data for a variety of natural language categories. Analysis of the categorization data for each of these categories yielded a single underlying scale along which both items and categorizers could be located (Verheyen et al., 2010). This suggests that a process that is similar to that proposed by the threshold theory underlies categorization decisions: Inter-individual differences in categorization can be shown to come about through the placement of threshold criteria along a latent scale that is common to all categorizers.

However, it does not indicate whether or not this scale reflects items' similarity to the category representation. This would require a substantive interpretation of the items' β estimates.

To provide such an interpretation, we adhere to the common assumption that features are the representational units of semantic concepts (e.g., O'Connor, Cree, & McRae, 2009; Storms, Navarro, & Lee, 2010) and that the similarity between an item and a category can be operationalized as the degree of feature overlap item and category display (e.g., Dry & Storms, 2010; Hampton, 1979). We would then expect a positive relationship between the number of features an item and a category have in common and the item's position β_i along the latent scale. This is akin to stating that the items that have the highest likelihood of being considered category members are the ones that have the most features in common with (i.e., are the highest in similarity to) the target category.

The original conception of the threshold theory would have us believe that only those features that are shared by item and category contribute to the likelihood that the item is endorsed as a category member. If we were to take the example of *avocados*, for instance, only those features that they have in common with the category of ***fruits***, like <*smell nice*> and <*taste fresh*>, would be thought to contribute to *avocados* being considered true ***fruits***. On the other hand, *avocados* embody features that distinguish them from many other ***fruits***. They <*are used in warm dishes*> and <*are nutritious*>. These are features that *avocados* share with the category of ***vegetables*** and might detract from *avocados*' likelihood of being considered ***fruits***. If this were to be the case, this would constitute evidence for a contrast category effect: Similarity toward a category other than the target category would impact negatively on the likelihood of categorization.

The emergence of contrast category effects would not be incompatible with the threshold theory's conception of the latent categorization scale. In fact, as Hampton already foresaw in his original exposition of the threshold theory, "increased emphasis on contrasting categories in the categorization task can easily be accommodated within a similarity based account" (Hampton, 1998, p. 157). If we were again to assume that feature overlap is a proxy measure of similarity, this accommodation would involve the inclusion of two such measures instead of a single one. The number of features item and target category have in common would be expected to yield a positive contribution to the item's position along the categorization scale. Thus, the higher an item's similarity toward the target category, the higher on the latent scale it would be expected to be located. The number of features the item has in common with a contrast category would be expected to yield a negative contribution to the item's position along the categorization scale. Thus, the higher an item's similarity toward a contrasting category, the lower on the scale it would be expected to be found. Both measures of feature overlap would have to contribute independently to establish a contrast category effect.

Within the framework of the probabilistic threshold model we can test whether both measures of feature overlap contribute. The linear logistic test model (Fischer, 1973, 1995) is a hierarchical extension of the Rasch model in that it expresses the β_is as a linear combination of item predictors. If we were to employ T_i to represent the number of features item i has in common with target category $\boldsymbol{T}$, and C_i to represent the number of features item i shares with contrast category $\boldsymbol{C}$, the resulting expression for β_i would read as follows:

$$\beta_i = \gamma_0 + \gamma_T T_i + \gamma_C C_i + \varepsilon_i \text{ with } \varepsilon_i \sim \mathrm{N}(0, \sigma_\varepsilon^2), \qquad (2)$$

with γ_T and γ_C expressing the effects of feature overlap with target and contrast, respectively, and γ_0 taking the role of intercept. To establish evidence in favor of contrast category effects both similarity measures would have to contribute independently to β_i. γ_T would have to be reliable and positive. At the same time γ_C would have to be reliable and negative.

It might be too strong of an assumption to argue that the items' βs can be perfectly predicted by their feature overlap with target and contrast. Features that are true of the items but do not signal overlap with target or contrast (i.e., distinctive features) might also contribute to the items' positions along the categorization scale (Tversky, 1977). It is also not clear to what extent other item variables, like familiarity or category dominance, exert an influence on the items' perceived category membership beyond that of the feature measures (Casey, 1992; De Deyne, 2008; Hampton, 1981, 1998; McCloskey, 1980). In Equation 2 we therefore allow random item variation to influence the prediction of the β_is. This not only has the advantage that the resulting model's fit is not invariably inferior to that of the Rasch model but also allows one to assess the magnitude of the explanatory power of the item predictors (De Boeck, 2008; Janssen, Schepers, & Peres, 2004). The inclusion of a random error term ε_i for every β_i releases the model from the requirement that the feature measures predict the item positions perfectly. σ_ε^2 then reflects the residual variance, just like in a regular regression model (Janssen, 2009).

Overview

In what follows we employ the extension of the probabilistic threshold model, expressed in Equation 2, to analyze various sets of categorization data. The data result from studies that differ in important methodological respects. This is fundamental, as it has been argued that whether a contrast category exerts its effect may depend on the specific task and instructions used (e.g., Goldstone, 1996; Markman & Makin, 1998; Ross, 1996; Yamauchi & Markman, 1998). To establish an influence of contrast categories on categorization we thus need to demonstrate their influence across a variety of conditions. For this reason we also employ different stimulus materials in different studies. What is common to all studies is that they involve superordinate, natural categories that adhere to the common definition of contrasting categories: They are mutually exclusive terms that are organized under an inclusive covering term (Goldstone, 1996; Martin & Billman, 1994; Verbeemen et al., 2001). All materials were presented in Dutch. All participants were fluent speakers of Dutch. The presented data are available from the first author's website at http://ppw.kuleuven.be/concat

All of the conducted studies will be described in a similar way. First, we describe the categorization task that yielded the dependent data. We then elaborate on the collection of feature data that allow for the computation of the measures of feature overlap that constitute the independent variables. These are then used in the analysis of the categorization data using the extension of the probabilistic threshold model. All analyses were performed in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). For every analysis five chains were run, each with a burn-in sample of 1,000 and 10,000 post-burn-in samples. The Markov Chain Monte Carlo methods implemented in WinBUGS were used specifically to sample the posterior distribution over the regression weights γ_T and γ_C. This allowed us to establish whether similarity to both the target category and a contrast category underlie categorization decisions. Similarity to the target category, operationalized as the amount of feature overlap between the items' and the target category's representation, is hypothesized to contribute positively to items' positions along the latent scale. Similarity to the contrast category, operationalized as the amount of feature overlap between the items' and the contrast category's representation, is hypothesized to contribute negatively to items' positions along the latent scale. For the scale to represent both types of similarity, the contributions of both item predictors should be reliable. In order for the categorization data to show a contrast category effect the mean of the posterior distribution over γ_T thus has to be positive, while the mean of the posterior distribution over γ_C has to be negative. In addition, the corresponding Bayesian credibility intervals (i.e., the region around the mean that contains 95% of the mass of the parameter's posterior distribution) may not include 0 (Gelman, Carlin, Stern, & Rubin, 2004).

Study 1: Carnivores and Herbivores

Material

In the overarching domain of animals the superordinate categories of *carnivores* and *herbivores* do not both apply to a particular animal (provided we leave ***omnivores*** aside). The two categories thus adhere to the definition of contrast categories introduced above. Verbeemen, Vanpaemel, Pattyn, Storms, and Verguts (2007) compiled a set of colored photographs of animals that could either be said to belong to the ***carnivore*** or to the ***herbivore*** category. An exemplar generation task informed the compilation of this set. Ten University of Leuven graduate students were asked to generate the names of 10 animals belonging to the category of ***carnivores*** and 10 animals belonging to the category of ***herbivores***. Photographs for all animals that were mentioned at least twice were included in the exemplar set. Thirty-five of these were ***carnivores***; 38 were ***herbivores***. Twenty-eight animals that were assumed to be rather unknown to students residing in Western Europe were included in the set as well, making for a total of 101 picture stimuli. A list with stimulus names is provided in Verbeemen et al.

Method

Categorization task. We constructed a categorization task in which participants were shown the 101 animal pictures in a random order. For each of the animal pictures they were asked to indicate whether the picture depicted a member of the ***carnivore*** category or of the ***herbivore*** category. Participants indicated their decision by clicking on one of two category labels that appeared next to each other below each picture. The presentation order of the category labels was constant for each individual participant but counterbalanced across participants. A total of 200 participants completed the categorization task.

Feature generation task. From Verbeemen et al. (2007) 39 features for the exemplar stimulus set were available. The procedure that they employed to obtain these features was very similar to the one used by Rosch and Mervis (1975) to elicit features for basic-level categories. One hundred and one participants were each presented with 10 of the well-known animal names and asked to generate up to 10 characteristic features for each of them. Each of the animal names was presented to 10 different participants. Those features that were generated at least 40 times across participants and stimuli were selected for use in a feature applicability task.

Feature applicability tasks. Verbeemen et al. (2007) organized the 101 animal photographs and the 39 animal features into an Excel file. This file contained 101 sheets, each holding one of the animal photographs as well as all of the animal features. Ten University of Leuven students indicated for each photograph–feature combination whether the feature applied to the animal depicted in the photograph or not, by entering a 1 or a 0 next to the feature in the Excel file. In case of uncertainty, participants were asked to provide their best guess. Participants performed the task at home and could freely choose when they worked on it. They were asked not to pause until all the features for a particular photograph were judged for applicability.

The features that are available from Verbeemen et al. (2007) were generated toward individual animal names. Because of this, information about which features are characteristic of ***carnivores*** and which are characteristic of ***herbivores*** is not readily available. We therefore conducted an additional feature applicability task in which one half of the participants was to indicate which of the 39 selected animal features they felt were characteristic of ***carnivores***. The other half was to indicate the features they felt were characteristic of ***herbivores***. They could do so by either putting down a 1 if they felt the feature generally applied to the category or putting down a 0 if they felt the feature did not generally apply to the category. Having separate groups of participants complete the task for the two categories instead of having participants choose whether a feature was either a ***carnivores*** or an ***herbivores*** one allowed features to be judged characteristic of both categories. This seemed to us a more likely circumstance for two categories pertaining to a single domain of entities. Two random orders of the features were presented to the participants. Each order–category combination was completed by 25 participants.

Analysis. The 10 item-by-feature applicability matrices from Verbeemen et al. (2007) were combined into a single matrix by determining the majority judgment for each cell. The same was done for the newly collected 2 × 50 category-by-feature applicability vectors. A feature was determined characteristic of a category when the majority of the participants decided this was the case. Two sets of analyses were then conducted with the extension of the probabilistic threshold model. In the first analysis the category of ***carnivores*** was considered the target category, while the category of ***herbivores*** was considered the potential contrast category. The dichotomized item-by-feature applicability matrix was used to determine how many of the features deemed characteristic of ***carnivores*** were true of each of the items. This resulted in a value T_i for every item. The number of ***herbivore*** features that were true of the items was determined to yield C_i. Note that in calculating C_i we took care not to incorporate those features that were deemed characteristic of both ***carnivores*** and ***herbivores*** so as not to mistake a contrast category effect for what is in fact an effect (partially) driven by similarity to the target category. T_i and C_i were standardized to have a mean of 0 and a standard deviation of 1. The second analysis was conducted in the same manner, but with the category of ***herbivores*** in the role of the target category and the category of ***carnivores*** as the potential contrast category.

Results

For each of the 101 items the proportion of participants who provided a positive categorization decision is depicted in Figure 2. The proportions are displayed as a function of the items' feature overlap with the target category (T_i) and of the items' feature overlap with the contrast category (C_i). In the left panel the category of ***carnivores*** is considered the target and the category of ***herbivores*** is considered the contrast. In the right panel the category of ***herbivores*** is considered the target and the category of ***carnivores*** is considered the contrast. Both panels demonstrate an increase in the proportion of positive categorization decisions with T_i and a decrease in the proportion of positive categorization decisions with C_i. The analyses of the individual categorization decisions with the extended threshold model indicate that both feature measures contribute reliably to the items' positions along the categorization scale in the directions suggested by the aggregated data in Figure 2.

With respect to the ***carnivore*** category the posterior mean of the target overlap weight γ_T was estimated at 1.5060 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.1860 and 1.8310, respectively. As 0 is not included in the credibility interval, this result signals a positive, reliable contribution of feature overlap with the target category to the items' positions along the latent scale. The more features items have in common with the target category, the higher their corresponding β_i estimates and the higher their overall probability of being endorsed as category members.

The posterior mean of the contrast overlap weight γ_C was estimated at −1.5149 with the 2.5% and 97.5% boundaries of the credibility interval equaling −1.8240 and −1.2060, respectively. As 0 is not included in the credibility interval, this result signals a negative, reliable contribution of feature overlap with the contrast category (***herbivores***) to the items' positions along the latent scale. The more features items have in common with the contrast category, the lower their corresponding β_i estimates and the lower their overall probability of being endorsed as category members.

When the same categorization data were analyzed with the category of ***herbivores*** considered the target category and the category of ***carnivores*** considered the contrast, a similar result was obtained. Feature overlap with target and contrast were found to yield independent and opposite contributions to the positioning of items along the latent scale. The posterior mean of the target overlap weight γ_T was estimated at 1.1885 with the 2.5% and 97.5% boundaries of the credibility interval equaling 0.9571 and 1.4190, respectively. As 0 is not included in the credibility interval, this result signals a positive, reliable contribution of feature overlap with the target category to the items' positions along the latent scale. The more features items have in common with the category of ***herbivores***, the higher their corresponding β_i estimates and the higher their overall probability of being endorsed as ***herbivores***.

The posterior mean of the contrast overlap weight γ_C was estimated at −1.9921 with the 2.5% and 97.5% boundaries of the credibility interval equaling −2.2420 and −1.74600, respectively. As 0 is not included in the credibility interval, this result signals a negative, reliable contribution of feature overlap with the contrast category (***carnivores***) to the items' positions along the latent scale. The more features items have in common with the category of ***carnivores***, the lower their overall probability of being endorsed as ***herbivores*** (i.e., the lower their β_i estimates).

Together the predictor variables account for 71% of the variance in the β_i estimates when the ***carnivore*** category is the target and 85% in the case the ***herbivore*** category is the target. This constitutes a considerable improvement over the contribution made by the target overlap predictor alone, which accounted for 43% of the variance in the former case and 37% in the latter.

Discussion

The results of both the analyses are in line with the idea that similarity underlies the organization of the latent categorization scale in the probabilistic threshold model. Items' similarity toward the target category was found to contribute positively to items' positions along the scale. The more similar they were to the target category, the higher their likelihood of being endorsed as a cate-

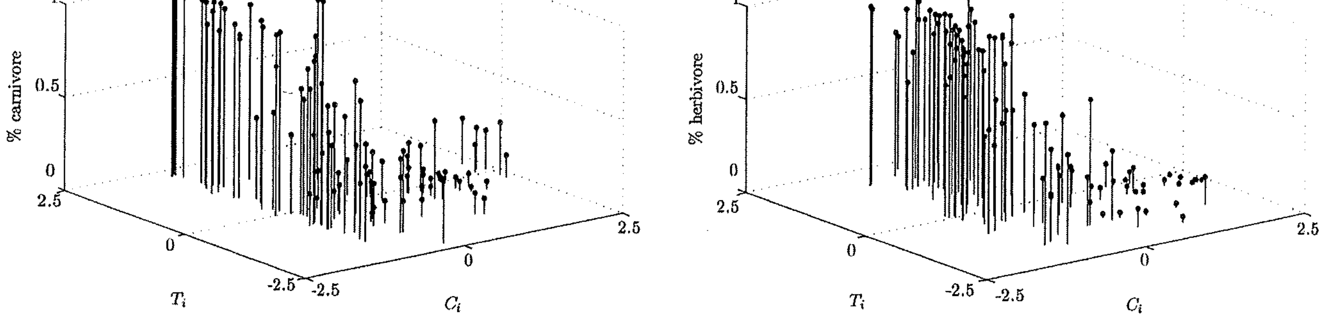

Figure 2. Categorization proportions as a function of feature overlap with the target category (T_i) and feature overlap with the contrast category (C_i) in Study 1.

gory member. This clearly provides support for the threshold theory's original conception of the categorization dimension as one that is representing similarity to the target category. In addition, the similarity of items toward a contrasting category was found to have a negative effect over and above that of items' similarity toward the target category. The more items resembled a contrasting category, the lower their likelihood of being included in the target category. This particular result suggests that the threshold theory's original conception of the categorization dimension needs to be extended to represent similarity to a contrast category in addition to similarity to the target category.

These results also establish a contrast category effect on categorization among natural, superordinate categories. To our knowledge, this study is the first to show such an effect for these kinds of categories. Study 1 demonstrates that it is possible for categories that share little representational structure (i.e., natural superordinate ones) to exert an influence on each other's extension.

The purpose of Study 2 was twofold. First, by employing plant stimuli instead of animal stimuli we wanted to generalize the finding obtained in Study 1 to a different set of natural categories. Second, we wanted to establish the contrast category effect using features that were generated toward the superordinate categories themselves instead of features that were generated toward the categories' exemplars, as was the case in Study 1. Features that are generated toward a category's exemplars provide for a richer representation than features that are generated toward the category (De Deyne, 2008). If one assumes that the absence of contrast category effects at the superordinate level in the literature is mainly due to lack of representational overlap, it should be more difficult to demonstrate a contrast category effect using the sparser superordinate category features. Study 2 thus constitutes a harder test for the existence of contrast category effects among natural, superordinate categories.

Study 2: Fruits and Vegetables

Material

The superordinate categories of ***fruits*** and ***vegetables*** together span a considerable range of the plant life that we grow for food. They are also generally thought of as contrasting categories. Storms, De Boeck, and Ruts (2001) employed the common procedure for establishing contrast categories to substantiate this claim. The 10 University of Leuven research assistants who were asked to imagine that they had heard a description of an entity and had provided ***fruit*** as a description all generated ***vegetable*** as a second guess. Nine out of the 10 research assistants who were told to imagine that their first guess had been ***vegetable*** provided ***fruit*** as a second guess. Smits, Storms, Rosseel, and De Boeck (2002) compiled a set of colored photographs of ***fruit*** and ***vegetable*** exemplars. An exemplar generation study by Storms, De Boeck, Van Mechelen, and Ruts (1996) informed the compilation of this set. Ten University of Leuven students were asked to generate up to 10 exemplars for the category of ***fruits***, while another 10 participants were asked to fulfill the same task for ***vegetables***. Photographs for exemplars that were generated at least twice were included in the set. Thirty-four of these were ***fruit*** exemplars, while 45 were ***vegetable*** exemplars. Smits et al. added 30 photographs of ***fruits*** and ***vegetables*** that were thought to be rather unknown to students residing in Western Europe. The final stimulus set thus consisted of 109 photographs. A list with stimulus names is provided in Smits et al.

Method

Categorization task. We constructed a categorization task that was similar to the one described in the first study. A total of 221 participants were shown the 109 plant pictures in a random order. For each of the plant pictures they were asked to indicate whether it depicted a member of the ***fruits*** category or of the ***vegetables*** category. Participants indicated their decision by clicking on one of two category labels that appeared next to each other below each picture. The presentation order of the category labels was constant for each participant but counterbalanced across participants.

Feature generation task. De Deyne et al. (2008) had participants generate up to 10 characteristic features of the categories of ***fruits*** and ***vegetables***. Twenty different participants generated features for each of the categories. None of the participants who generated features for ***fruits*** generated features for ***vegetables*** and vice versa. We selected all features that were generated by at least two participants for inclusion in a feature applicability task. A total of 32 such features were generated toward ***fruits***; a total of 29 different features were generated toward ***vegetables***. Due to overlap between the features generated for the two categories, the total number of features used in the applicability task equaled 53.

Feature applicability task. To obtain applicability judgments we employed the procedure taken by Verbeemen et al. (2007). The 109 plant photographs were placed on separate sheets of an Excel file along with the 53 features that were obtained in the generation task. Eight paid volunteers were asked to work their way through the file, placing a 1 next to each feature they felt applied to the plants depicted in the photographs and placing a 0 next to the features they felt did not apply to the plants depicted in the photographs. Participants performed the task at home and could freely choose when they worked on it. They were, however, asked not to pause until all the features for a particular photograph were judged for applicability.

Note that because the features were generated toward the two categories involved in the categorization task instead of toward the categories' exemplars, it is not necessary to include an additional applicability task to determine the category representations. The features that were generated toward a category label can be taken to make up that category's representation.

Analysis. The eight participants' applicability matrices were combined into a single matrix by determining the majority judgment for each cell. To obtain a measure of T_i we determined for each item the number of features that were generated toward the category that was considered the target and at the same time were true of the item in the dichotomized applicability matrix. The number of features that were both generated toward the category acting as contrast and true of the item was determined to yield C_i. In calculating C_i we took care not to incorporate those features that were generated toward both ***fruit*** and ***vegetable***. T_i and C_i were both standardized to have a mean of 0 and a standard deviation of 1. As in Study 1, two sets of analyses were conducted with the extension of the probabilistic threshold model: one with the category of ***fruits*** considered the target and the category of ***vegetables***

considered the potential contrast and another with the category of ***vegetables*** considered the target and the category of ***fruits*** considered the potential contrast.

Results

For each of the 109 items the proportion of participants who provided a positive categorization decision is depicted in Figure 3. The proportions are displayed as a function of the items' feature overlap with the target category (T_i) and of the items' feature overlap with the contrast category (C_i). In the left panel the category of ***fruits*** is considered the target and the category of ***vegetables*** is considered the contrast. In the right panel the category of ***vegetables*** is considered the target and the category of ***fruits*** is considered the contrast. Both panels demonstrate an increase in the proportion of positive categorization decisions with T_i and a decrease in the proportion of positive categorization decisions with C_i. The analyses of the individual categorization decisions with the extended threshold model indicate that both feature measures contribute reliably to the items' positions along the categorization scale in the directions suggested by the aggregated data in Figure 3.

When the ***fruits*** category was considered the target, the posterior mean of the target overlap weight γ_T was estimated at 2.9888 with the 2.5% and 97.5% boundaries of the credibility interval equaling 2.6190 and 3.3710, respectively. As 0 is not included in the credibility interval, this result signals a positive, reliable contribution of feature overlap with the target category to the items' positions along the latent scale. The more features that items have in common with the target category, the higher their corresponding β_i estimates and the higher their overall probability of being endorsed as category members.

In the same analysis the posterior mean of the contrast overlap weight γ_C was estimated at –2.6037 with the 2.5% and 97.5% boundaries of the credibility interval equaling –3.0110 and –2.2110, respectively. This result signals a negative, reliable contribution of feature overlap with the contrast category (***vegetables***) to the items' positions along the latent scale. The more features items have in common with the contrast category, the lower their corresponding β_i estimates and the lower their overall probability of being endorsed as category members. The inclusion of feature overlap with the contrast category constitutes an increase in the variance of β_i accounted for from 41% to 81%.

When the same categorization data were analyzed with the category of ***vegetables*** considered the target category and the category of ***fruits*** considered a potential contrast category, a similar result was obtained. Feature overlap with target and contrast were found to yield independent and opposite contributions to the positioning of items along the latent categorization scale. The posterior mean of the target overlap weight γ_T was estimated at 1.8980 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.5260 and 2.2840, respectively. As 0 is not included in the credibility interval, this result again signals a positive, reliable contribution of feature overlap with the target category to the items' positions along the scale. The more features that items have in common with the ***vegetables*** category, the higher their overall probability of being endorsed as ***vegetables***.

The posterior mean of the contrast overlap weight γ_C was estimated at –3.3172 with the 2.5% and 97.5% boundaries of the credibility interval equaling –3.7050 and –2.9430, respectively. This again signals a negative, reliable contribution of feature overlap with the contrast category (***fruits***) to the items' positions along the latent scale. The more features items have in common with the category of ***fruits***, the lower their corresponding β_i estimates and the lower their overall probability of being endorsed as ***vegetables***. The inclusion of feature overlap with the contrast category constitutes an increase in the variance of β_i accounted for from 15% to 81%.

Discussion

In Study 2 the contrast category effect that was found in the first study was generalized to another set of natural superordinate categories. Similarity toward a contrast category proved to exert an influence on categorization decisions over and above that of similarity to the target, and this occurred despite the use of superordinate category features to determine similarity. These results again confirm the threshold theory's claim that similarity drives categorization decisions, with the understanding that similarity needs to be determined with respect to both target and contrast category.

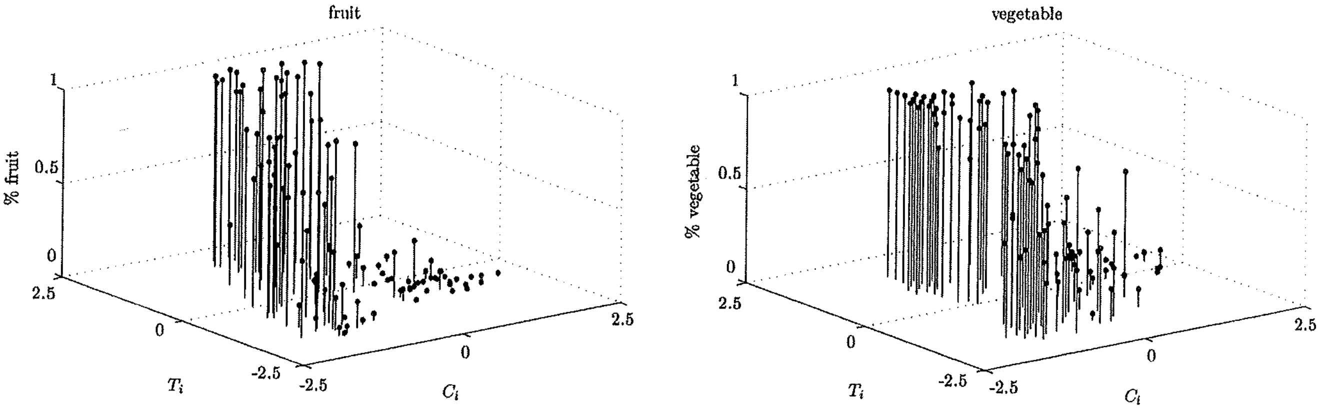

Figure 3. Categorization proportions as a function of feature overlap with the target category (T_i) and feature overlap with the contrast category (C_i) in Study 2.

One might object to the claim that the contrast category effect observed in both studies constitutes evidence for the representation of the categories as each other's contrasts. The observed effect might merely reflect the setup of the studies in which participants were confronted with a choice between two categories (Davis & Love, 2010; Goldstone, 1996). The main purpose of Study 3 was therefore to establish whether contrast category effects of the kind that were found in Studies 1 and 2 can also be shown when the contrast category is not presented during categorization. To this end, we analyze categorization data that were taken from Verheyen et al. (2010) in which participants were simply told to indicate whether or not items belonged to a target category. The dependent data thus constitute *yes* and *no* decisions instead of choices between two categories.

The stimulus material that was employed by Verheyen et al. (2010) comes with another advantage. While the items that were presented to the participants in Studies 1 and 2 were clearly sampled from two categories (the target category and the hypothesized contrast category), the items in Verheyen et al. were not. They were Dutch translations of the items employed by Hampton, Dubois, and Yeh (2006) to study context effects on categorization. The stimulus lists Hampton et al. compiled for each category were composed of true members, borderline members, and nonmembers. The latter items were not chosen with a particular contrast for the target category in mind. This makes it unlikely that the nonmembers included in the stimulus lists cue participants into encoding the various items in terms of the particular contrasts we employ (Medin & Edelson, 1988). If contrast category effects were still to arise under these circumstances, they are probably due to the categories' being represented as each other's contrasts.

A final point of difference between Study 3 and the previous studies is that verbal items instead of pictorial ones were employed. In processing pictorial rather than verbal stimuli participants have been shown to emphasize information that distinguishes a particular item from others (Gati & Tversky, 1984; Ritov, Gati, & Tversky, 1990). This suggests that the effect of features shared with a contrast category, which set an item apart from other potential members, might be more difficult to obtain using verbal items.

Study 3 is similar to Studies 1 and 2 in that it includes plant and animal categories situated at the superordinate level of abstraction. We also continue to use features generated toward the superordinate category labels to determine T_i and C_i.

Study 3: Fruits, Vegetables, Fish, and Insects

Material

Hampton et al. (2006) included four natural superordinate categories with 24 items each, which were translated into Dutch by Verheyen et al. (2010). The categories can be organized into two domains that span two of the superordinate categories each. The categories of ***fruits*** and ***vegetables*** constitute plants that are cultivated by humans for food. The categories of ***fish*** and ***insects*** belong to the domain of animals. These pairs of categories thus satisfy the requirements to be termed contrast categories: They are mutually exclusive terms that are organized under an inclusive covering term. A list of the items is provided in Hampton et al.

Method

Categorization task.

The categorization data were taken from Verheyen et al. (2010), who had 250 University of Leuven students provide category membership judgments for the items included under ***fruits***, ***vegetables***, ***fish***, and ***insects***. Participants also provided categorization decisions for four additional, nonnatural superordinate categories. They could indicate whether or not they felt a particular item belonged to the target category by circling 1 for *yes* or 0 for *no*. They were also given the opportunity to indicate that a particular item was unknown to them. Five different orders of category administration were combined with two different orders of item administration, resulting in 10 different questionnaires. Each of these was filled out by 25 participants.

Feature generation task. For the four superordinate categories involved in the categorization task, characteristic features were available from De Deyne et al. (2008). The features were obtained by having 80 University of Leuven students (20 for each category) list those features they felt were important for something to be considered a member of the category. Participants were asked to generate up to 10 of these features. Only features generated toward a category by at least two participants were selected. For the categories of ***fruits*** and ***vegetables*** this resulted in 32 and 29 features, respectively. These are the same features that were used in Study 2. For the category of ***fish*** 31 features were selected. Thirty-five features were selected for the category of ***insects***.

Feature applicability task. We constructed two item-by-feature matrices, one for each of the two domains covered in the categorization study. The 2 × 24 items of the corresponding domains made up the rows of the matrices, while the features that were deemed characteristic of the two categories in each domain made up the columns. Due to overlap between the features generated for the categories, the total number of features used in the applicability task for the plant items equaled 53. It totaled 63 in the applicability task for the animal items. Five University of Leuven students indicated for each item–feature pair in a matrix whether the feature applied to the item or not by entering a 1 or a 0 in the corresponding matrix cell. In case of uncertainty, the participants were asked to provide their best guess. 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.

As in Study 2, the features were generated toward the category labels, and therefore there was no need to include an additional applicability task to determine the category representations. The features that were generated toward a category label were taken to make up that category's representation.

Analysis. The item-by-feature matrices for the plant domain were combined into a single matrix by determining the majority judgment for each cell. So were the item-by-feature matrices for the animal domain. For the 24 items included in the plant applicability matrix that were categorized with respect to ***fruit***, we determined the number of features in common with the ***fruit*** representation to obtain a measure of T_i. To obtain a measure of C_i for these items we determined the number of features they had in common with the ***vegetable*** representation. Within the same applicability matrix we repeated this procedure for the 24 items that were categorized with respect to ***vegetable***, and with ***fruit*** and ***vegetable***, respectively, taking the roles of contrast and target. The 2×24 items in the animal applicability matrix were subjected to the same procedure, involving the representations of ***fish*** and ***insect***. Each time C_i was calculated, we took care not to incorporate those features that were generated toward both categories that made up a matrix. The resulting T_i and C_i measures were each standardized to have a mean of 0 and a standard deviation of 1.

Four sets of analyses were then conducted with the extension of the probabilistic threshold model: one with the category of ***fruits*** considered the target and the category of ***vegetables*** considered the potential contrast; a second with the category of ***vegetables*** considered the target and the category of ***fruits*** considered the potential contrast; a third with the category of ***fish*** considered the target and the category of ***insects*** considered the potential contrast; and a fourth with the category of ***insects*** considered the target and the category of ***fish*** considered the potential contrast. Unlike the analyses conducted for the first two studies, the current analyses pertain to independent data sets with different items.

Results

The left upper panel of Figure 4 displays for each of 24 items the proportion of participants who endorsed it as a ***fruit***. The proportions are displayed as a function of the items' feature overlap with the category of ***fruits*** (T_i) and of the items' feature overlap with the category of ***vegetables*** (C_i). The right upper panel of Figure 4 displays for each of 24 items the proportion of participants who endorsed it as a ***vegetable***. The proportions are displayed as a function of the items' feature overlap with the category of ***vegetables*** (T_i) and of the items' feature overlap with the category of ***fruits*** (C_i). The left lower panel of Figure 4 displays for each of 24 items the proportion of participants who endorsed it as a ***fish***. The proportions are displayed as a function of the items' feature overlap with the category of ***fish*** (T_i) and of the items' feature overlap with the category of ***insects*** (C_i). The right lower panel of Figure 4 displays for each of 24 items the proportion of participants who endorsed it as an ***insect***. The proportions are displayed as a function of the items' feature overlap with the category of ***insects*** (T_i) and of the items' feature overlap with the category of ***fish*** (C_i). All four panels demonstrate an increase in the proportion of positive categorization decisions with T_i and a decrease in the proportion of positive categorization decisions with C_i. The analyses of the individual categorization decisions with the extended threshold model indicate that both feature measures contribute reliably to the items' positions along the categorization scale in the directions suggested by the aggregated data in Figure 4.

With regards to the contribution of feature overlap with the target to the items' positions along the categorization scale, the results are in line with those found in Studies 1 and 2. Feature commonality with the target category was found to have a reliable, positive contribution to the items' positions in that the more features they share with the target category, the higher their β_i. This was found for the category of ***fruits*** for which the posterior mean of the target overlap weight γ_T was estimated at 2.5333 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.7670 and 3.3280; the ***vegetables*** category with the posterior mean of the target overlap weight estimated at 2.4124 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.2720 and 3.5560; the category of ***fish*** for which the posterior mean of the target overlap weight was estimated at 2.5694 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.6930 and 3.5080; and the ***insects*** category for which the posterior mean of the target overlap weight was estimated at 1.8135 with the 2.5% and 97.5% boundaries of the credibility interval

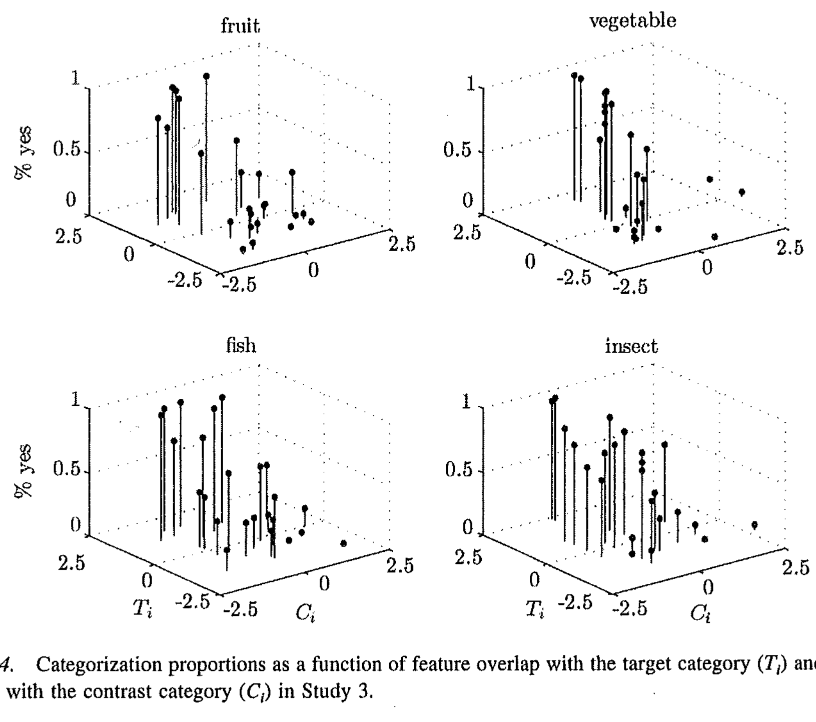

Figure 4. Categorization proportions as a function of feature overlap with the target category (T_i) and feature overlap with the contrast category (C_i) in Study 3.

equaling 1.1060 and 2.5480. As 0 is not included in any of these credibility intervals, these results signal a positive, reliable contribution of feature overlap with the target category to the items' positions along the categorization scale. The more features items have in common with the target category, the higher their corresponding β_i estimates and the higher their overall probability of being endorsed as category members.

The results concerning the contribution of feature overlap with a contrast, too, correspond to the findings obtained in the first two studies. The posterior mean of the contrast overlap weight γ_C was estimated at –1.5086 for the ***fruits*** category; at –1.1459 for ***vegetables***; at –1.0765 for ***fish***; and at –0.9076 for ***insects***. None of the corresponding credibility intervals included 0. The 2.5% and 97.5% boundaries of the credibility interval equaled –2.3170 and –0.7352 for ***fruits***; –2.2750 and –0.0093 for ***vegetables***; –1.9770 and –0.1832 for ***fish***; and –1.6370 and –0.1908 for ***insects***. These results signal a negative, reliable contribution of feature overlap with a contrast category (respectively, ***vegetables***, ***fruits***, ***insects***, and ***fish***) to the items' positions along the latent scale. The more features items have in common with the contrast category, the lower their corresponding β_i estimates and the lower their overall probability of being endorsed as category members.

In the case of ***fruits*** the inclusion of the contrast category predictor increased the variance of β_i accounted for from 60% when only target category information was included to 77%. In the case of ***vegetables*** there was an increase from 41% to 50%, from 64% to 72% in the case of ***fish***, and from 59% to 68% in the case of ***insects***.

Discussion

It has been argued that the existence of contrast category effects is task and instruction dependent (Goldstone, 1996). One effective manner in which it can be achieved is through an informed choice of the basis of comparison in a categorization task (Davis & Love, 2010). In Study 3 we ensured that there was no such basis of comparison prominent. Participants were simply asked to decide whether items belonged in the target category or not (instead of having to categorize items into one of two contrasting categories), and the nonmembers did not constitute cues to the categories we chose as contrasts. In addition, the stimuli were verbal items and the categories natural and situated at the superordinate level. Despite these circumstances, a clear contrast category effect emerged: Categorization in each of the target categories was influenced both by similarity to the target and by similarity to a contrast.

An attentive reviewer remarked that the inclusion of more than one target category in the categorization task could potentially suffice to have the categories behave as each other's contrast. The subsequent presentation of the target categories might then result in a contrast effect of categories that were presented early in the task on categories that were presented later in the task. The design of the categorization task, which included five different orders of category administration, allows for a test of the reviewer's hypothesis. We can test whether the contrast category effect is still reliable when we restrict our analyses to the data pertaining to categories that haven't been preceded by the hypothesized contrast category (e.g., analyze the ***fruit*** categorization decisions from the participants who made ***fruit*** categorization decisions before they made ***vegetable*** categorization decisions). For ***fruits*** the posterior mean of γ_T was established at 2.5370 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.7690 and 3.3600, respectively. The posterior mean of γ_C was established at –1.4607 with the 2.5% and 97.5% boundaries of the credibility interval equaling –2.2460 and –0.7033, respectively. For ***vegetables*** the posterior mean of γ_T was established at 2.4445 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.2790 and 3.6370, respectively. The posterior mean of γ_C was established at –1.3273 with the 2.5% and 97.5% boundaries of the credibility interval equaling –2.4920 and –0.1390, respectively. For ***fish*** the posterior mean of γ_T was established at 2.5114 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.6800 and 3.4780, respectively. The posterior mean of γ_C was established at –1.2232 with the 2.5% and 97.5% boundaries of the credibility interval equaling –2.1020 and –0.4045, respectively. For ***insects*** the posterior mean of γ_T was established at 1.9530 with the 2.5% and 97.5% boundaries of the credibility interval equaling 1.2260 and 2.7090, respectively. The posterior mean of γ_C was established at –0.9483 with the 2.5% and 97.5% boundaries of the credibility interval equaling –1.6850 and –0.2148, respectively. Thus, the contrast category effect remained reliable in these restricted data sets, supporting our earlier findings and rejecting the claim that these might have resulted from the subsequent presentation of categories.

The influence of similarity to an implicit contrast category on items' positions along the latent scale in the probabilistic threshold model strongly suggests that the categories involved are represented as each other's contrasts. This in turn warrants the incorporation of contrast information in the threshold theory framework in particular and categorization models for natural language categories in general.

General Discussion

The three main contributions of the current work lie in (a) the establishment of contrast category effects in categorization among natural superordinate categories, (b) the introduction of a data analytic tool for categorization data, and (c) its implications for the threshold theory and other accounts of categorization. Below we elaborate on each of these matters.

Contrast Category Effects

Studies 1–3 demonstrate that it is possible for natural superordinate categories from the same domain to exert an influence on each other's extension. Using a variety of stimulus materials and tasks we showed that the decision to endorse an item as a member of a natural superordinate category can be informed by both the item's similarity to the target category and its similarity to an alternative category from the same domain. This result extends previous research that had already shown contrast category effects on categorization for natural language categories with considerable representational overlap (i.e., artifact superordinate categories; Verbeemen et al., 2001). It suggests that contrast category effects on categorization might be more pervasive than has previously been thought. In fact, the results reported here might even constitute an underestimate of the contrast category effect. After all, there is no principled reason why only one contrasting category should exert an influence on the target category. It is very likely that multiple alternative categories exert an influence on categorization in a target category. Evidence for this claim comes from the common procedure for eliciting contrasts, which generally yields more than one alternative to the suggested target category. Of the 86 participants in Verbeemen et al. (2001) who were subjected to this procedure, only 67 generated ***fruits*** in response to ***vegetables*** and 60 generated ***vegetables*** in response to ***fruits***. Even if one takes into account that some of the generated contrasts will not reside at the superordinate level, this leaves ample opportunity for other categories to emerge as potential contrasts. The category of ***herbs***, for instance, is a plausible candidate to fulfill the role of contrast category for ***fruits*** and ***vegetables*** in Studies 2 and 3. Other animal categories might function as contrasts for the ***fish*** and ***insects*** categories in Study 3. It would be straightforward to test for an influence of multiple contrast categories within the framework that we have proposed. One would have to determine only the feature overlap between the experimental stimuli and the potential contrast categories to test for an influence of these categories on categorization in the target category.

When applied in this manner, the extension of the threshold model can be employed to determine which categories can be considered contrasts for a category under scrutiny. Admittedly, the suggested procedure to determine contrasts is an explorative one, as our contribution remains agnostic as to when items' similarity to an alternative category from the same domain exerts an influence on categorization in a particular target category. The presented studies merely show that it is possible to find contrasting categories for natural superordinate categories. We do not contend that all potential alternatives will prove to be contrasts. It is not unlikely that one could establish a considerable degree of similarity between a set of experimental stimuli and a category, without showing a contrast category effect of this category. Clearly, future work should be directed at establishing the necessary conditions for contrast category effects. The modeling framework that we forwarded in this article can easily be brought to bear on the issue of the origins of contrast category effects. The model can be employed as a data analytic tool to test hypotheses regarding the matter, or it can be extended to include variables that might be responsible for the emergence of the contrast category effects. Whether such manipulations affect the contribution of a potential contrast category to categorization then becomes apparent in the value of the γ_C parameter that expresses the effect of feature overlap with the contrast. In the next section we provide an example of both these approaches using a hypothesis regarding contrast category effects that was put forward by Verbeemen, Storms, and Verguts (2003). It states that as categories become more familiar, the influence of contrasting categories will disappear (see also Corneille, Goldstone, Queller, & Potter, 2006).

Data Analytic Tool

Studies 1 and 2 allow for a test of the familiarity hypothesis, as they included a number of items that were considered rather unknown to our participants. If lack of familiarity with these items can be thought responsible for the emergence of the contrast category effect in the two studies, the effect should disappear when the unfamiliar items are excluded from the analyses. Application of the extended probabilistic threshold model to all but the 28 unfamiliar items in Study 1 yielded reliable contributions of γ_T to β_i for both ***carnivores*** and ***herbivores***. For ***carnivores*** the posterior mean of the target overlap weight was estimated to be 1.6472 with the corresponding credibility interval bounded between 1.2280 and 2.0720. It was estimated to be 1.1990 with the corresponding credibility interval bounded between 0.9301 and 1.4730 for ***herbivores***. The amount of feature overlap with the contrast category yielded a reliable negative contribution to β_i both in the case of ***carnivores*** and in the case of ***herbivores***. For the category of ***carnivores*** the posterior mean of the contrast overlap weight was estimated to be –1.6377 with the corresponding credibility interval bounded between –2.0220 and –1.2570. For the category of ***herbivores*** the posterior mean of the contrast overlap weight was estimated to be –2.2393 with the corresponding credibility interval bounded between –2.5490 and –1.9360. Contrary to the familiarity hypothesis, the credibility intervals for γ_C did not include 0 when the unfamiliar items were removed from the analyses.

Exclusion of the novel items did not affect the results obtained for Study 2 either. Feature overlap with target and contrast continued to exert independent, reliable contributions to the items' positions along the categorization scale. Excluding the 30 stimuli that were thought to be unknown to our participants yielded a γ_T posterior mean estimate of 3.2839 and a γ_C posterior mean estimate of –2.7290 for ***fruit***. The corresponding credibility intervals were bounded between 2.8080 and 3.7780 and between –3.3000 and –2.1830, respectively. The posterior means for γ_T and γ_C were estimated at 2.2958 and –3.5053, respectively, for the category of ***vegetables***. The corresponding credibility intervals were bounded between 1.7190 and 2.9000 and between –3.9610 and –3.0710. None of the credibility intervals obtained in the reanalysis of the Study 2 data included 0.

The setup of Studies 1 and 2, with their inclusion of stimuli that were deemed unfamiliar to participants, is such that it allows for a test of a specific hypothesis. The extended probabilistic threshold model allows this test to be executed. In this case, the analyses using the model did not support the familiarity hypothesis.

Verheyen and Storms (2011) obtained familiarity ratings for the items included in the third study. They had 40 University of Leuven students rate each of the items on a scale ranging from 1 (*never seen, heard, or used*) to 7 (*seen, heard, or used very often*). These ratings offer an additional opportunity to asses the familiarity hypothesis. It is straightforward to include the familiarity ratings as an additional predictor in Equation 2. If unfamiliarity were to be responsible for the emergence of the contrast category effect in Study 3, the inclusion of the familiarity predictor should eliminate the contribution of the contrast overlap measure. This proved the case only for the category of ***vegetables*** where the credibility interval for γ_C now included 0. The ***vegetables***' posterior mean for γ_C was estimated to be –0.7584 with the credibility interval ranging from –1.5950 to 0.1183. The contribution of feature overlap with the contrasting category remained reliable and negative for the other three superordinates. The category of ***vegetables*** was also the only one for which the contribution of familiarity proved reliable. The posterior mean of the familiarity regression weight was estimated at –1.8171 with the credibility interval ranging from –2.7720 to –0.9470. Taken together with the reliable contribution of feature overlap with the target category, this result suggests that similarity with the target by itself might sometimes yield an overestimate of the categorization likelihood. The effect of familiarity might reflect an adjustment of this likelihood through participants' biological knowledge (i.e., knowing that an item is not a ***vegetable*** despite its similarity to the category) or it might constitute a correction of participants' confidence in endorsing features for familiar rather than unfamiliar items. Whatever the eventual explanation of the disappearance of the contrast category effect for ***vegetables***, it is clear that the current results do not provide much support for the familiarity hypothesis. For the majority of categories the contrast category effect on categorization persists even after familiarity is accounted for.

Bringing external empirical measures into the model to test whether they can account for the variability found among the model's parameters sets the current work apart from the application of the probabilistic threshold model by Verheyen et al. (2010), where it was merely fit to empirical categorization data to establish that it provides a satisfying account of the generated response patterns. In the current application we attempted to explain the variability that was found among items. In a similar manner external variables can be brought into the model to elucidate the sources of variability among persons (Van den Noortgate & Paek, 2004). One could, for instance, imagine verifying whether a context manipulation induces a change in the employed threshold criterion. Interactions between variables on the item and person side of the model can also be introduced (Meulders & Xie, 2004). One could, for instance, test whether persons making categorization decisions in one context place a different emphasis on target and contrast similarity than persons making categorization decisions in another context. This would allow one to tie the current work in with that of Goldstone (1996) and Davis and Love (2010) on the influence of task manipulations on the emergence of contrast category effects, and with that of Gati and Tversky on the varying importance of different types of feature information for different stimulus materials (Gati & Tversky, 1984; Ritov et al., 1990). Although we have here restricted the use of the model to the question of whether contrast category effects can be shown among natural superordinate categories, it constitutes in fact an exceedingly flexible tool for the analysis of categorization data.

Threshold Theory

The prominence of the contrast category effect in the current studies might appear somewhat surprising in light of its virtual absence in the categorization literature relating to natural language categories. We believe that our ability to demonstrate these effects is due to our use of a model that is specifically tailored toward membership decisions pertaining to natural language categories. The model we employed is specifically intended to capture the inter-individual categorization differences that result from these categories' vague boundaries. As mentioned in the introduction, the analyses employed in Verbeemen et al. (2001) provided very little evidence in favor of contrast category effects. The reason for this may be due to their use of regression analyses to explain the variability in categorization decision times. Vandekerckhove, Verheyen, and Tuerlinckx (2010) have recently demonstrated that various sources of variability may be confounded in such analyses. Further, they demonstrated that it is possible to disentangle the different variability sources using a model that represents the different processes that underlie speeded categorization decisions. The model has the added benefit that it allows for the simultaneous analysis of the categorization decisions and the speed with which they were made. If a similar process model approach were applied to the Verbeemen et al. data then it is possible that they may also yield contrast category effects. Of course, it is also possible that the processing constraints associated with having to respond within a specified time window do not allow for the computation of both similarity to the target and similarity to a contrast.

Establishing contrast category effects has proven less difficult with regards to typicality, presumably because the employed accounts are flexible enough to allow for influences other than those of the category under scrutiny. The contributions by Ameel and Storms (2006) and by Voorspoels et al. (2008) provided clear examples of this. The work by Ameel and Storms pertains to the prediction of typicality from items' similarity to a prototype in a geometrical representation. Instead of determining the position of the prototype by taking the average of the positions of the category members, Ameel and Storms allowed the prototype to move freely in the representational space. The prototype that ended up best predicting typicality was not located among the category members but lay further away in a direction away from the other categories represented in the space. Several natural superordinate categories were included among the categories that showed this contrast category effect (see Voorspoels et al., 2008, for a related illustration involving the categories of ***birds***, ***fish***, ***insects***, ***mammals***, and ***reptiles***).

With these models the extended probabilistic threshold model shares the idea that the most representative items are the ones with the highest similarity to the target category *and* the lowest similarity to the contrast. The model shows closest resemblance, however, to an account of typicality that was introduced by Dry and Storms (2010). Dry and Storms used the contrast model (Tversky, 1977) to generalize the polymorphous concept measure (Hampton, 1979). Instead of assuming that only the features that items have in common with the target category or its exemplars contribute to the items' perceived typicality, Dry and Storms showed how features that are distinctive to the items also contribute. When these distinctive features signal overlap with alternative categories from the same domain, these contributions can be thought of as contrast category effects (Verheyen et al., 2011). The model we introduced in this article can thus be regarded an extension of the generalized polymorphous concept account. It too highlights the importance of feature overlap both with the target category and with contrasting categories but proposes a mechanism that allows these measures to be related to binary categorization decisions.

Note that the relationship with the work of Tversky (1977) also implies that the model is easily extended to incorporate effects of features that are distinctive but do not signal overlap with a contrast category. This can be illustrated for the data from Study 1 in which a number of the features were not deemed characteristic of the ***carnivore*** or the ***herbivore*** category. Hence they do not signal overlap with the hypothesized contrast when they apply. (It is not possible to show this using the data from Studies 2 and 3, as these include only features that are characteristic of the categories involved.) The number of these distinctive features D_i that apply to an item i can be incorporated in Equation 2 along with T_i and C_i. The contribution of D_i was found positive and reliable for ***carnivore*** (mean γ_D is 0.5280 with the credibility interval bounded between 0.2334 and 0.8220). It was found negative and not reliable for ***herbivore*** (mean γ_D is -0.1724 with the credibility interval bounded between −0.4060 and 0.0620). Incorporating distinctive feature effects in the analyses does not alter the earlier conclusions. The posterior mean of γ_T was established at 1.5877 when ***carnivore*** was the target category. Its credibility interval was bounded between 1.2830 and 1.8990. In the same analysis the posterior mean of γ_C was established at −1.4401 with the credibility interval bounded between −1.7360 and −1.1430. When ***herbivore*** was the target category, the posterior mean of γ_T was established at 1.1410 with its credibility interval bounded between 0.9038 and 1.3760. The corresponding posterior mean of γ_C was established at −1.9987 with its credibility interval bounded between −2.2490 and −1.7560. The contrast category effect clearly remains after the incorporation of distinctive features. For both ***carnivores*** and ***herbivores*** the contribution of T_i was still found positive and reliable, while the contribution of C_i remained negative and reliable.

In the original conception of the threshold theory (Hampton, 1998, 2007) typicality is taken to be a proxy of similarity to the target category. In Verheyen et al. (2010) a strong positive correlation between the probabilistic threshold model's β_is and rated typicality was therefore taken to support the idea that the model's latent scale represents similarity to the category representation. This particular result, taken together with the results from Studies 1–3 and those from (among others) Ameel and Storms (2006) and Voorspoels et al. (2008), suggests that this interpretation of typicality requires adjusting. The typicality ratings, too, are suspected to be influenced by both the category toward which the ratings are made and a potential contrasting category. In order to verify this we obtained typicality ratings for the stimuli that were used in Studies 1, 2, and 3 and employed the generalized polymorphous concept account (Dry & Storms, 2010) to look for an influence of both target feature overlap and contrast feature overlap on typicality.

The typicality ratings were obtained from Vanpaemel, Ameel, and Storms (2011) and Verheyen et al. (2010), who provided participants with a scale from 1 (*very atypical*) to 7 (*very typical*) to indicate their judgments. Vanpaemel et al. presented the 101 photographs of novel and well-known animals from Study 1 to 52 University of Leuven students with the question to rate their typicality either as a carnivore or as an herbivore. Twenty-seven participants rated the animals' typicality toward ***carnivores***. Twenty-five participants rated the animals' typicality toward ***herbivores***. They also presented the 109 photographs of novel and well-known fruits and vegetables from Study 2 to two independent groups of judges. Twenty-nine participants rated each depicted plant's typicality toward ***fruit***. Twenty-one participants rated each depicted plant's typicality toward ***vegetables***. Forty participants in Verheyen et al. rated the 24 items in each of the four superordinate categories for their typicality. The correlation between rated typicality and items' estimated positions β_i along the categorization scale in the probabilistic threshold model was higher than .90 in all eight categories, demonstrating a close relationship between items' representativeness and their likelihood of being endorsed as category members.

For each of the eight target categories the generalized polymorphous concept account (Dry & Storms, 2010) was implemented. Like the extended probabilistic threshold model, the generalized polymorphous concept account allows for a contribution of feature overlap with the target category (T_i) and of feature overlap with the contrast category (C_i). The values of T_i and C_i that were used in Studies 1–3 were also used here. The generalized polymorphous concept (GPC) account includes a single parameter ρ to determine the relative influence of T_i and C_i on typicality:

$$GPC_i = \rho T_i - (1 - \rho)C_i. \quad (3)$$

If ρ equals 1 the generalized polymorphous concept measure includes only features that items and the target category have in common. If ρ equals 0 the generalized polymorphous concept measure includes only features that items have in common with the contrast category. Values of ρ that lie in between 0 and 1 signal a contribution of both types of similarity. We varied the value of ρ from 0 to 1 in increments of .0005 and for every value computed the correlation between the resulting generalized polymorphous concept measure and rated typicality. We report the ρ values that yielded the optimal correlation between the generalized polymorphous concept measure and rated typicality. With respect to typicality toward the category of ***carnivores***, a ρ estimate of .44 was obtained. With respect to the typicality ratings toward the category of ***herbivores***, ρ was estimated to equal .55. Analysis of the typicality ratings for the plant pictures with the generalized polymorphous concept account yielded a ρ of .55 for ***fruit*** and .47 for ***vegetables***. In the case of the four superordinate categories ***fruits***, ***vegetables***, ***fish***, and ***insects***, the analysis of typicality resulted in ρ estimates of .50, .68, .59, and .45, respectively. These ρ estimates all deviate considerably from 1, suggesting that both similarity to target and similarity to contrast exert an influence on the typicality ratings.

These observations are supported by the corresponding increases in the correlation with the typicality ratings. In the case of ***carnivores*** the optimal correlation between the generalized polymorphous concept account and typicality is established at .84, compared with .74 when only feature overlap with the target category is considered. Including feature overlap with a contrast category increases the correlation with typicality from .70 to .90 in the case of ***herbivores***, from .80 to .91 in the case of the pictured ***fruits***, and from .52 to .83 in the case of the pictured ***vegetables***. Increases from .83 to .89, from .70 to .79, from .86 to .88, and from .79 to .83 were observed for the remaining categories of ***fruits***, ***vegetables***, ***fish***, and ***insects***. In all eight categories the Akaike information criterion (Akaike, 1973) favors the account with a contribution of both feature overlap with the target category and feature overlap with the contrast category over the account that incorporates only feature overlap with the target category.

It thus appears that when participants are asked to judge an item's typicality for a particular target category, they make use of information that is relevant for distinguishing the item from (implicit) contrast categories. Indeed, unlike the categorization tasks in Studies 1 and 2 where participants were required to choose between two categories, the typicality rating task for the ***carnivores***, ***herbivores***, ***fruits***, and ***vegetables*** categories mentioned only one category. Of course, participants could have inferred the relevant contrast category through their exposure to items that are clear members of the contrast category. That criticism, however, does not apply to the typicality ratings for the four other superordinate categories. The nonexemplars in the Verheyen et al. (2010) study were not informative with respect to the contrast categories we employed in the current study. Categories' representations might thus well include information about contrasting categories.

The ramifications of this are most evident with regard to the threshold theory of categorization (Hampton, 1998, 2007). We used a formalization of the theory to analyze categorization data and found that the underlying categorization dimension represents not only similarity to the target category but also similarity to a contrast category. While similarity to the target category was found to contribute positively to items being perceived as members, the similarity of items to an alternative category detracted from their perceived membership in the target category. The same conclusion was reached when the items' rated typicality, a strong correlate of the items' positions along the categorization dimension, was analyzed using the generalized polymorphous concept account (Dry & Storms, 2010). As Hampton (1998) already foresaw, these contrast category effects did not prove at odds with the threshold theory. Within its similarity framework the effect was easily accommodated for through the inclusion of a measure of items' feature overlap with the contrast category among the predictors of the items' positions along the categorization dimension. In this article we put the threshold theory of categorization to test. However, contrast category effects pose a challenge to all accounts of categorization. As the empirical evidence for contrast category effects accumulates, the challenge for future researchers is to find ways to incorporate the influence of both target and contrast category information within the categorization processes they propose.

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.

Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language, 55,* 402–421. doi:10.1016/j.jml.2006.05.005

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37,* 372–400. doi:10.1006/jmps.1993.1023

Casey, P. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 823–834. doi:10.1037/0278-7393.18.4.823

Ceulemans, E., & Storms, G. (2010). Detecting intra- and inter-categorical structure in semantic concepts using HICLAS. *Acta Psychologica, 133,* 296–304. doi:10.1016/j.actpsy.2009.11.011

Corneille, O., Goldstone, R. L., Queller, S., & Potter, T. (2006). Asymmetries in categorization, perceptual discrimination, and visual search for reference and nonreference exemplars. *Memory & Cognition, 34,* 556–567. doi:10.3758/BF03193579

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science, 21,* 234–242. doi:10.1177/0956797609357712

De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73,* 533–559. doi:10.1007/s11336-008-9092-x

De Deyne, S. (2008). *Proximity in semantic vector space* (Unpublished doctoral dissertation). University of Leuven, Leuven, Belgium.

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. doi:10.3758/BRM.40.4.1030

Dry, M. J., & Storms, G. (2010). Features of graded category structure: Generalizing the family resemblance and polymorphous concept models. *Acta Psychologica, 133,* 244–255. doi:10.1016/j.actpsy.2009.12.005

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,* 359–374. doi:10.1016/0001-6918(73)90003-6

Fischer, G. H. (1995). The linear logistic test model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 131–155). New York, NY: Springer.

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16,* 341–370. doi:10.1016/0010-0285(84)90013-6

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London, England: Chapman & Hall/CRC.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24,* 608–628. doi:10.3758/BF03201087

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition, 31,* 169–180. doi:10.3758/BF03194377

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18,* 441–461. doi:10.1016/S0022-5371(79)90246-9

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition, 9,* 149–156. doi:10.3758/BF03202329

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65,* 137–165. doi:10.1016/S0010-0277(97)00042-5

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31,* 355–384. doi:10.1080/15326900701326402

Hampton, J. A., Dubois, D., & Yeh, W. (2006). Effects of classification context on categorization in natural categories. *Memory & Cognition, 34,* 1431–1443. doi:10.3758/BF03195908

Hampton, J. A., Estes, Z., & Simmons, C. L. (2005). Comparison and contrast in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1459–1476. doi:10.1037/0278-7393.31.6.1459

Janssen, R. (2009). Modeling the effect of item designs within the Rasch model. In S. Embretson & J. Roberts (Eds.), *New directions in psychological measurement with model-based approaches* (pp. 227–245). Washington, DC: American Psychological Association.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189–212). New York, NY: Springer.

Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44. doi:10.1037/0033-295X.99.1.22

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York, NY: Wiley.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337. doi:10.1023/A:1008929526011

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language, 29,* 289–315. doi:10.1016/0749-596X(90)90002-H

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language, 31,* 195–217. doi:10.1016/0749-596X(92)90011-L

Malt, B. C., & Sloman, S. A. (2004). Conversation and convention: Enduring influences on name choice for common objects. *Memory & Cognition, 32,* 1346–1354. doi:10.3758/BF03206325

Markman, A., & Makin, V. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127,* 331–354. doi:10.1037/0096-3445.127.4.331

Markman, A., & Wisniewski, E. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 54–70. doi:10.1037/0278-7393.23.1.54

Martin, J. D., & Billman, D. O. (1994). Acquiring and combining overlapping concepts. *Machine Learning, 16,* 121–155. doi:10.1007/BF00993176

McCloskey, M. E. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior, 19,* 485–502. doi:10.1016/S0022-5371(80)90330-8

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6,* 462–472. doi:10.3758/BF03197480

Medin, D., & Edelson, S. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85. doi:10.1037/0096-3445.117.1.68

Meulders, M., & Xie, Y. (2004). Person-by-item predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 213–240). New York, NY: Springer.

Nosofsky, R. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114. doi:10.1037/0278-7393.10.1.104

O'Connor, C., Cree, G., & McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science, 33,* 665–708. doi:10.1111/j.1551-6709.2009.01024.x

Palmeri, T. J., & Nosofsky, R. M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology, 54*(A), 197–235. doi:10.1080/02724980042000084

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danish Institute for Educational Research.

Ritov, I., Gati, I., & Tversky, A. (1990). Differential weighting of common and distinctive components. *Journal of Experimental Psychology: General, 119,* 30–41. doi:10.1037/0096-3445.119.1.30

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605. doi:10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. doi:10.1016/0010-0285(76)90013-X

Ross, B. H. (1996). Category representations and the effects of interacting with instances. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1249–1265. doi:10.1037/0278-7393.22.5.1249

Ruts, W., Storms, G., & Hampton, J. A. (2004). Linear separability in superordinate natural language concepts. *Memory & Cognition, 32,* 83–95. doi:10.3758/BF03195822

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin & Review, 9,* 836–844. doi:10.3758/BF03196343

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of novel stimuli in well-known natural concepts: A case study. *Psychonomic Bulletin & Review, 8,* 377–384. doi:10.3758/BF03196176

Storms, G., De Boeck, P., Van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1–15. doi:10.1037/0278-7393.22.5.1266

Storms, G., Navarro, D. J., & Lee, M. D. (2010). Introduction to the special issue on formal modeling of semantic concepts. *Acta Psychologica, 133,* 213–215. doi:10.1016/j.actpsy.2009.11.004

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51,* 567–577. doi:10.1007/BF02295596

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi:10.1037/0033-295X.84.4.327

Vandekerckhove, J., Verheyen, S., & Tuerlinckx, F. (2010). A crossed random effects diffusion model for speeded semantic categorization decisions. *Acta Psychologica, 133,* 269–282. doi:10.1016/j.actpsy.2009.10.009

Van den Noortgate, W., & Paek, I. (2004). Person regression models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 167–187). New York, NY: Springer.

Vanpaemel, W., Ameel, E., & Storms, G. (2011). *Is prototypical typical? An investigation using semantic concepts.* Manuscript submitted for publication.

Verbeemen, T., Storms, G., & Verguts, T. (2003). Determinants of speeded categorization in natural concepts. *Psychologica Belgica, 43,* 139–151.

Verbeemen, T., Vanoverberghe, V., Storms, G., & Ruts, W. (2001). The role of contrast categories in natural language concepts. *Journal of Memory and Language, 44,* 618–643. doi:10.1006/jmla.2000.2748

Verbeemen, T., Vanpaemel, W., Pattyn, S., Storms, G., & Verguts, T. (2007). Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach. *Journal of Memory and Language, 56,* 537–554. doi:10.1016/j.jml.2006.09.006

Verheyen, S., Hampton, J. A., & Storms, G. (2010). A probabilistic threshold model: Analyzing semantic categorization data with the Rasch model. *Acta Psychologica, 135,* 216–225. doi:10.1016/j.actpsy.2010.07.002

Verheyen, S., & Storms, G. (2011). Does a single dimension govern categorization in natural language categories? In B. Kokinov, A. Karmiloff-Smith, & N. J. Nersessian (Eds.), *European perspectives on cognitive science.* Sofia, Bulgaria: New Bulgarian University Press.

Verheyen, S., Stukken, L., De Deyne, S., Dry, M. J., & Storms, G. (2011). The generalized polymorphous concept account of graded structure in abstract categories. *Memory & Cognition.* Advance online publication. doi:10.3758/s13421-011-0083-2

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Modeling typicality: Extending the prototype view. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 751–756). Austin, TX: Cognitive Science Society.

Voorspoels, W., Vanpaemel, W., & Storms, G. (in press). Contrast in natural language concepts: An exemplar-based approach. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

Yamauchi, T., & Markman, A. (1998). Category learning by inference and categorization. *Journal of Memory and Language, 39,* 124–148. doi:10.1006/jmla.1998.2566

Received July 23, 2010
Revision received May 13, 2011
Accepted May 16, 2011 ■