This article was downloaded by: [K.U.Leuven - Tijdschriften]
On: 30 April 2012, At: 08:57
Publisher: Psychology Press
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pqje20

Contrast effects in typicality judgements: A hierarchical Bayesian approach

Wouter Voorspoels [a], Gert Storms [a] & Wolf Vanpaemel [a]

[a] Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium

Available online: 01 Feb 2012

To cite this article: Wouter Voorspoels, Gert Storms & Wolf Vanpaemel (2012): Contrast effects in typicality judgements: A hierarchical Bayesian approach, The Quarterly Journal of Experimental Psychology, DOI:10.1080/17470218.2012.662237

To link to this article: http://dx.doi.org/10.1080/17470218.2012.662237



PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2012, i*First*, 1–19

Psychology Press
Taylor & Francis Group

Contrast effects in typicality judgements: A hierarchical Bayesian approach

Wouter Voorspoels, Gert Storms, and Wolf Vanpaemel

Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

We examine the influence of contrast categories on the internal graded membership structure of everyday concepts using computational models proposed in the artificial category learning tradition. In particular, the generalized context model (Nosofsky, 1986), which assumes that only members of a given category contribute to the typicality of a category member, is contrasted to the similarity–dissimilarity generalized context model (SD-GCM; Stewart & Brown, 2005), which assumes that members of other categories are also influential in determining typicality. The models are compared in a hierarchical Bayesian framework in their account of the typicality gradient of five animal categories and six artefact categories. For each target category, we consider all possible relevant contrast categories. Three separate issue are examined: (a) whether contrast effects can be found, (b) which categories are responsible for these effects, and (c) whether more than one category influences the typicality. Results indicate that the internal category structure is codetermined by dissimilarity towards potential contrast categories. In most cases, only a single contrast category contributed to the typicality. The present findings suggest that contrast effects might be more widespread than has previously been assumed. Further, they stress the importance of characteristics particular of everyday concepts, which require careful consideration when applying computational models of representation of the artificial category learning tradition to everyday concepts.

Keywords: Concepts; Categories; Typicality; Contrast effects; Computational modelling of behavioural data.

Cows are mammals. They are good, representative members of the class of mammals, displaying many characteristics that are common in this class: They live on land, give birth to live young, have nipples, breath air, have fur, and so on. The same cannot be said for all members of the class of *mammals*. The platypus, for instance, is a rather bad mammal, lacking many of the common properties: While it does have fur and breathe air, it is also semiaquatic, lays eggs, and does not have nipples. Such differences in representativeness of members of a category, often referred to as the typicality gradient of a category, can be reliably observed in a broad range of everyday language concepts (e.g.,

Correspondence should be addressed to Wouter Voorspoels, K.U. Leuven, Department of Psychology, ConCat, Tiensestraat 102, B-3000, Leuven, Belgium. E-mail: wouter.voorspoels@psy.kuleuven.be

The research in this article is part of research project G.0281.06 sponsored by the Belgian National Science Foundation–Flanders, given to the second author. We thank Steven Verheyen and Sander Vanhaute for help with collecting data and for fruitful input, and Tim Bosschaerts for programming the sorting task. Website: ppw.kuleuven.be/concat

© 2012 The Experimental Psychology Society

http://www.psypress.com/qjep http://dx.doi.org/10.1080/17470218.2012.662237

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

Barsalou, 1983; De Deyne et al., 2008; Rosch & Mervis, 1975). Moreover, typicality has been shown to relate to a robust processing advantage across a number of category-based tasks (e.g., Keller & Kellas, 1978; Malt & Smith, 1984; Rips, 1975, Rosch, 1975; for an overview see Hampton, 1993), making it one of the most studied variables in concept research.

The typicality gradient is often thought of as reflecting the *internal* membership structure of a concept—that is to say, it is a characteristic of one concept, independent of other concepts. In this perspective, one would expect that it relies solely on the representation of the concept and derives, for example, from the similarity to other members of a concept's extension (e.g., Brooks, 1978), or relates to the similarity to the prototype of the category (e.g., Posner & Keele, 1968). This intuition has been the basis for a number of models, which have been shown fairly successful in their account of the typicality gradient of everyday concepts (e.g., Hampton, 1979; Rosch & Mervis, 1975).

An important characteristic of everyday concepts, however, is that they are not isolated entities, but reside in rich semantic networks that provide important sources of meaning (e.g., Bechtel, 1998; Block, 1999; Collins & Quillian, 1969; Dietrich & Markman, 2003; McNamara & Miller, 1989; Millikan, 1984; Landauer & Dumais 1997). A dictionary illustrates this by providing meaning to the concepts it lists through reference to other concepts and the conceptual network they are part of. For example, people less familiar with the platypus might benefit from knowing that it has a beak, resembling that of *ducks*, it has poison glands as do *insects* and *reptiles*, it lays eggs similar to *birds* and *reptiles*, and is semiaquatic as are *amphibians*. These concepts all relate to animals, and to each other, and are part of a semantic network.

The embeddedness of concepts in semantic networks elicits an intriguing question regarding graded membership in concepts: Is the internal typicality gradient of a concept determined solely by characteristics internal to the category, or do neighbouring concepts influence a concept's meaning? When considering membership of an object in the extension of a concept, do people only consider the concept at hand, or do they also consider whether the object resembles the members of another concept's extension? If it resembles a member of another category, this information is quite informative for deciding whether it is a member of the target concept. In terms of typicality, the question then is whether the platypus is an atypical member of mammals merely because of its apparent dissimilarity towards other mammals, or does its obvious similarity to other animal categories, such as *birds* and *reptiles*, contribute to its atypicality? In what follows, we reserve the term *contrast effect* for this kind of influence of a concept on another concept's internal membership structure.

Outline

The present paper examines the extent to which neighbouring categories contribute to typicality. We focus on everyday natural language concepts but use computational models of typicality that have their roots in artificial category learning research. In particular, we compare two exemplar-based accounts of typicality. The generalized context model (GCM; Nosofsky, 1986) assumes that typicality of an exemplar in a category is a function of the summed similarity to all other category members. The idea that an exemplar's typicality derives from both the similarity to category members and the dissimilarity to members of alternative categories is implemented in the similarity–dissimilarity GCM (SD-GCM; Stewart & Brown, 2005)—a variant of the GCM.

The models are unified in a hierarchical Bayesian framework. The framework serves as a way to relate models to the observed data and allows making inference on interesting latent variables that govern the model predictions (Lee, 2011). The framework elegantly allows addressing of three fundamental questions regarding contrast effects in natural language categories: First, is there evidence for contrast effects—that is, the influence of a concept on another concept's internal membership structure? If this is the case, inferences can be made to parameters that embody the two additional questions regarding the specific contrast categories that exert the contrast effects: which of

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

the contrast categories exert contrast effect, and how many do so.

We start with providing some background on contrast effects. In particular we highlight the particular challenges faced when attempting to examine contrast effects in everyday language concepts by applying computational models from the artificial category learning tradition. Next, we present the data, consisting of two semantic domains, animals and artefacts, that present a good test of contrast effects. The two basic models, the GCM and the SD-GCM, are then presented separately, followed by a description of the hierarchical Bayesian approach taken to relate models and data. In this approach, the two models are unified, and inferences to the key questions, translated to parameters in the framework, can be made. Next, the results of the hierarchical Bayesian analyses are presented and discussed.

Contrast effects and contrast categories

A contrast effect occurs when a category-based judgement not only derives from similarity with the target category, but also takes into account dissimilarity with an alternative category. In artificial categorization, the existence of contrast effects are well established (e.g., Davis & Love, 2010; Goldstone, 1996; Goldstone, Steyvers, & Rogovsky, 2003; Palmeri & Nosofsky, 2001; Stewart & Morin, 2007). For example, Palmeri and Nosofsky (2001) demonstrated that items residing at the target's category boundary most remote from other categories were classified most accurately. In other words, clear category members were both high in similarity towards the target category and low in similarity towards other relevant categories. Goldstone et al. (2003) manipulated the interrelatedness of categories and demonstrated a clear contrast effect when categories are sufficiently interrelated.

While the findings in artificial category learning strongly suggest that contrast effects in related concepts exist, this does not easily generalize to everyday concepts, since they do not replicate the specific conditions that apply to everyday concepts. Everyday concepts reside in a much richer network, containing more than one alternative category. For example, the semantic domain of *living things* contains a myriad of concepts, each with a unique extension. Contrary to the predefined categories in artificial category learning, it is not justified to decide a priori which, if any, of these categories exert contrast effects on a target category.

In everyday concepts the question thus not only is whether typicality in concepts can be influenced by other concepts, but also which concepts exert influence on a target concept. It is plausible to assume that the conceptual interrelatedness is an important factor in determining candidate concepts (Goldstone, 1996). For example, whereas one might plausibly hypothesize that the concepts *fish* and *bird* can exert contrast effects on each other, it is rather implausible that the concepts *justice* and *doorknob* do so. A particularly interesting type of related concepts are *contrast categories*. Contrast categories are considered to be at the same level of abstraction as a target category, belonging to the same immediate superordinate. Further, they are mutually exclusive in the sense that one and only one word is applicable to any member of the category (Miller & Johnson-Laird, 1976). For example, *bird* and *fish* are contrast categories, both belonging to the same immediate superordinate category of animals, and they are perfectly mutually exclusive (an animal cannot be a bird and a fish at the same time). Contrast categories are obviously very closely related, yet quite separate, and therefore particularly relevant when deciding, for example, whether a newly encountered animal is a bird.

While the focus on contrast categories narrows down the set of candidates considerably, most everyday language categories will have multiple related categories that adhere to the definition of contrast category. The category *bird* has four obvious contrast categories: *mammals*, *fish*, *insects*, and *reptiles*, each of which might exert contrast effects on the typicality gradient of *fish*. Obviously, there is no a priori reason why a concept would have only one contrast category: *Fish* might very well have more than one contrast category exerting influence—for example, *mammals* and *birds*. Thus, when moving to natural language concepts, the possibility of multiple categories exerting contrast effects needs careful consideration (Rosch & Mervis, 1975; Verbeemen, Vanoverberghe, Storms, & Ruts, 2001).

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

A second difference with the studies mentioned above is that the paradigms applied in artificial category learning generally consist of discriminating between two categories, possibly facilitating contrast effects. Deciding whether a stimulus is in Category A implies that it is not in B, or the other way around (see Davis & Love, 2010, for a different approach). Judgements of typicality, which are the focus in the present research, do not have the explicit reference to a concept other than the target concept.

In sum, when moving from artificial categories to the realm of everyday concepts, the issue of contrast effects covers a broader load than only their existence. The first and perhaps most important question remains whether contrast effects come into play when people judge the typicality of members of everyday concepts. However, even if we narrow the search for candidate concepts to the set of contrast categories of a concept, two additional issues are raised: which of the candidate concepts exerts influence on a target concept's typicality gradient, and, related to this question, how many of the candidates do so.

Data

The present study focuses on two semantic domains that lend themselves perfectly to the examination of contrast effects: the animal and artefact domain. Both domains contain a number of superordinate concepts that adhere nicely to the definition of contrast categories and can thus be considered candidates to exert contrast effects on each other. For example, the animal domain contains superordinate categories that are at the same level of abstraction, share a common superordinate (animals), and are mutually exclusive. For each of these categories, we can thus evaluate whether the other categories—that is, contrast categories—exert influence on their typicality gradient. Moreover, we can examine which of the other categories exerts contrast effects, without a priori restricting the set of candidates, and we can investigate whether more than one does so. The same holds for the artefact domain.

In total, we consider 11 everyday concepts from the two semantic domains (De Deyne et al., 2008). The animal domain contains five categories (*mammals*, *birds*, *fish*, *insects*, and *reptiles*), each category containing between 22 and 30 exemplars, for a total of 129 exemplars. The artefact domain consists of six categories: *clothes, kitchen utensils, musical instruments, tools, vehicles, and weapons*. The domains consist of 166 exemplars, each category containing between 20 and 30 exemplars.

The focus of the present study is typicality of the exemplars of the categories, which is assessed using a goodness-of-example task. Further, the computational models of typicality are based on underlying geometric stimulus representations. This geometric stimulus representation for both domains is constructed based on a similarity measure derived from feature data.

Typicality

De Deyne et al. (2008) asked 28 participants to rate the exemplars of each category for goodness-of-example for the category to which they belonged. A 20-point Likert-rating scale was used ranging from 1 for very bad examples to 20 for very good examples. The reliability of the judgements for each category was evaluated by means of split-half correlations corrected with the Spearman–Brown formula. The estimated reliabilities ranged from .91 to .98. In the present study, the typicality ratings are averaged across participants, resulting in a single typicality score for each member of a category.

Stimulus representation

The computational models applied in the present context are built on underlying geometric stimulus representations. In a geometric stimulus representation, stimuli are represented as points in an *M*-dimensional space, where the distance between two points is inversely related to the similarity between the two stimuli. To allow the SD-GCM to include information from contrast categories, it is necessary that all the exemplars of a domain are in the same stimulus representation. Deriving an appropriate representation requires a measure of pairwise similarity between all exemplars of each

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

domain. On the basis of the pairwise similarity, geometric stimulus representations were built for the two semantic domains.

Pairwise similarity. De Deyne et al. (2008) report two large exemplar by feature matrices, one for the animal domain and one for the artefact domain. Each exemplar by feature matrix contains all exemplars of the domain and all features generated at least four times for its exemplars. The animal domain matrix contains 129 exemplars and 765 features; the artefact domain contains 166 exemplars and 1,294 features. The features were generated for the exemplars and span a broad range of characteristics. In total, 1,003 participants participated in the generation task. For each exemplar, at least 20 participants were asked to generate 10 features,[1] and a minimum of 180 features were generated for each exemplar.

For each domain separately, four participants judged the applicability of each feature for each exemplar (1 referring to applicable, 0 referring to not applicable), resulting in four feature-by-exemplar applicability matrices for each domain, with zero and one entries. The estimated reliability of the applicability judgements was .83 for the animal domain and .81 for the artefact domain. For each domain, the four feature-by-exemplar applicability matrices were then summed across participants. The summed matrix contains a feature applicability vector for each exemplar of the domain (values ranging from 0 to 4), reflecting how applicable each feature is for the exemplar. A pairwise measure of similarity for each pair of exemplars in a domain was derived by calculating the correlation between their corresponding feature applicability vectors.

Similarity scaling. To generate geometric similarity representations for the two domains, the pairwise similarity measure for each pair of exemplars of a domain was used as input in a SAS nonmetric multidimensional scaling (MDS) analysis (SAS, V9). The appropriateness of the geometric representation was evaluated using stress, a badness-of-fit measure. Following Kruskal and Wish (1978), we opted for solutions with a stress-value lower than .1. For both domains, this was the case from Dimensionality 4 onwards. For our model analyses, we will rely on the stimulus representation of Dimensionality 4.

Models

Both the GCM and the SD-GCM are exemplar based, which means that the models assume that people rely on stored exemplars to account for category-based decision. For natural language categories, exemplars are generally assumed to be defined at the nearest subordinate member level (e.g., Komatsu, 1992). For example, the exemplars of the category *mammals* are *dog*, *camel*, *beaver*, and so on.

Generalized context model

The GCM is one of the most influential exemplar models in categorization research, and it has been successfully applied to both categorization decisions (Nosofsky, 1986) and typicality judgements (Nosofsky, 1988; Voorspoels, Vanpaemel, & Storms, 2008). Typicality of an exemplar i for category A is assumed to be related to the similarity of that exemplar to all other exemplars in the category:

$$t_{iA} = \sum_{j \in A} s_{ij}$$

where s_{ij} is the similarity of exemplar i to exemplar j. A crucial assumption of the GCM is that similarity is not invariant but is presumed to change in constrained and systematic ways depending on the particular context. In the model, this assumption is implemented by allowing selective attention to play a central role in determining the similarity (Nosofsky, 1984, 1986). In other words, the similarity between two exemplars can vary according to the differential weight that is attributed to the dimensions that span the geometric stimulus representation. Selective attention is one of the

[1] Not all participants succeeded in generating 10 features.

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

crucial mechanisms behind the successful performance of the GCM, as well as other models of categorization. The similarity between two exemplars is a function of the distance of the exemplars in the psychological space:

$$S_{ij} = exp(-cd_{ij})$$

with d_{ij} the distance of exemplars i and j and c the sensitivity parameter. The distance is derived from the position of the exemplars in an M-dimensional geometric stimulus representation. Crucially, the distance can vary across contexts by differentially weighting the underlying stimulus dimensions. Formally, the weighted distance is given by:

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

where x_{ik} and x_{jk} are the coordinates of exemplars i and j on dimension k, and w_k is the dimension weight for dimension k. We relied on Euclidean distances (i.e., r is fixed at 2), which are generally accepted to be more appropriate for integral dimensions (Shepard, 1987). The free parameters in the GCM as applied in the present study consist of $M - 1$ dimension weights w_k (because the weights are constrained to sum to 1) and the sensitivity parameter c.

Similarity–dissimilarity GCM

The SD-GCM (Stewart & Brown, 2005) is an extension of the GCM, making identical assumptions concerning underlying representation and similarity. Whereas the GCM considers only similarity towards members of the category, the SD-GCM also takes into account the dissimilarity to members of alternative categories. The model has been originally proposed in the context of categorization decisions, but can be adapted to account for typicality. In the terminology of Stewart and Brown (2005), typicality of a particular exemplar for a category can be considered the evidence that the exemplar is a member of the category. The evidence for a particular category is a function of the summed similarity towards category members and the summed dissimilarity to members of the alternative category. Formally, the typicality of exemplar i to category A then is given by:

$$t_{i,A} = \sum_{j \in A} S_{ij} + \sum_{l \notin A} (1 - S_{il})$$

where s_{ij} are similarities of exemplar i to exemplars of category A, and s_{il} represent similarities towards exemplars of other categories. Typicality to a category is thus assumed to be related to the sum of the similarity towards members of the category as well as the summed dissimilarity to exemplars of other categories in the same representational space. The calculation of similarity is analogous to that of the GCM, making use of the same free parameters.

In the context of artificial categorization, the alternative category contributing to the evidence that an exemplar belongs to a target category is a priori and quite straightforwardly defined. In the context of natural language categories, however, the question is not only whether contrast effects exist, but which categories exert such an influence on the target category's internal structure. In other words, not all alternative categories in the representational space are necessarily included in the calculation of the dissimilarity component.

A Bayesian hierarchical approach

The SD-GCM, as applied in the present study, does not specify a priori which categories contribute to the dissimilarity component. One reasonable assumption would be that all contrast categories are involved, but it is also sensible to restrict the set of contrast categories, such that only a subset of exemplars in the representational space is included. For example, in accounting for the typicality of members of the category *mammals*, we might choose to restrict the set of contrast exemplars to members of the categories *birds* and *fish*, leaving out the other contrast categories *insects* and *reptiles*. In this way, several version or instantiations of the SD-GCM are possible. Inclusion of a particular contrast set—that is, a certain

combination of contrast categories—defines a particular instantiation of the SD-GCM, so in any given semantic domain the SD-GCM therefore counts multiple instantiations. The fact that the SD-GCM has multiple instantiations (depending on the specific contrast set included), while the GCM has only one (no contrast set is included), obviously makes the SD-GCM a more complex model than the GCM (e.g., Myung, 2000), which has only one instantiation—no inclusion of contrast categories—although each particular instantiation of the SD-GCM is probably roughly as complex as the GCM.

Overall, we have a large family of model instantiations at our disposal, ranging from a version where no contrast categories play a role (the GCM), to a model where all relevant categories contribute to typicality, as well as several models in between, where a subset of categories contributes useful information (all versions of the SD-GCM). A large family of closely related models is most conveniently analysed in a hierarchical framework (Lee, 2011; Vanpaemel, 2011). This approach, which involves specifying a generative mechanism for all model instantiations under consideration, is taken in the present study. The key strengths of the hierarchical Bayesian approach relevant for the present context include the ability to make inferences about parameters that correspond directly to the questions asked and the ability to elegantly deal with the vast complexity differences between the GCM and the SD-GCM. Both issues are discussed below.

The hierarchical extension

Each of the model instantiations in the framework has three salient and informative characteristics, respectively referring to whether a contrast set is included, how large the contrast set is, and which categories are in it. Using a hierarchical Bayesian framework, these characteristics can be quantified in hyperparameters. Figure 1 provides a schematic representation of the present hierarchical analysis, showing three levels.

The bottom level represents the observed typicality ratings—that is, the average across participants for each member of the target category. At the middle level, the target category exemplar coordinates and the coordinates of the exemplars of the selected contrast set can be found. On the basis of these coordinates, similarities between the relevant members are computed by means of the dimensional weights w_k and the sensitivity parameter c, as detailed above. This process is identical for both models under consideration, the GCM and the SD-GCM. The combination of the similarity towards target category exemplars and the dissimilarity towards contrast category exemplar determines the typicality score of exemplar i.

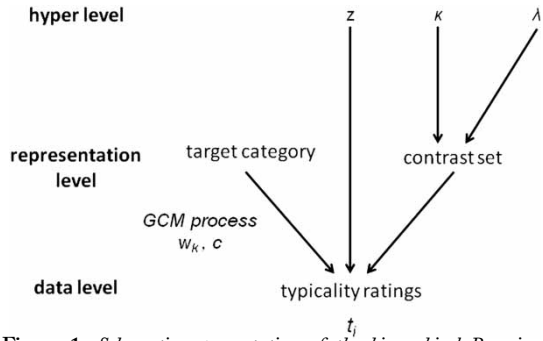


Figure 1. *Schematic representation of the hierarchical Bayesian framework.*

The top level corresponds to the hierarchical extension that unifies the GCM and SD-GCM in a single framework. The extension involves a simple mechanism to generate all model instantiations, which is governed by three parameters: z, κ, and λ. First, z decides whether contrast information is used in the calculation of the typicality scores. The parameter can have value 1 or 0, respectively indicating that the dissimilarity towards contrast category members is included in the typicality calculation or that the dissimilarity component is not used in the calculation. Second, the κ parameter controls how many categories are included in the contrast set. The parameter can take on natural values ranging from 1 to the maximum number of potential contrast categories. For example, if κ is three, the contrast set consists of three categories. Finally, the λ parameter governs the particular constitution of the contrast set.

Together, these three parameters determine the probability of a specific model instantiation being

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

generated. For example, if z is fixed at 0, no contrast categories are used, so the GCM representation results. If, in contrast, z is set at 1, a contrast set is added to the calculation. Two additional parameters then need consideration to generate a model instantiation: κ and λ, respectively determining the size and composition of the contrast set. When κ is 4, all alternative animal categories for a particular animal category are in the contrast set; when κ is set at 1, only one contrast category is in the contrast set; the specific composition of this contrast set—that is, which categories are part of the set—is determined by λ. For κ is 1, λ can be set at each of the alternative categories; if κ is 2, λ can specify each of the possible combinations of two alternative categories.

Once a model instantiation has been generated, the model predictions follow from the (SD-)GCM process. Figure 2 shows, for fixed values of w and c, how these predictions differ for the three hyper parameter values discussed above for the *mammals* category. In the figure, for each stimulus (represented on the x-axis, and rank ordered according to the model-based score), the predicted typicality score is shown (solid line) and is compared to the observed typicality score (cross). The outer panels in Figure 2 present the generated typicality scores when no (left-hand panel) or all (right-hand panel) contrast sets are used. The middle panel shows the predictions when only one contrast category is in the contrast set (*birds*).

Priors

A priori, we consider each contrast size as equally probable, implying a uniform prior for κ. Further, we assume that all potential contrast categories are equally probable, implying a uniform prior for λ. The prior for z is more tricky, and more crucial, as it is used to put the GCM and SD-GCM at roughly equal footing in terms of complexity. If all model instantiations (i.e., all possible contrast sets, ranging from no contrast categories to all contrast categories) would be given an equal prior mass, the hypothesis that contrast effects play a role would be a priori favoured over the hypothesis that contrast effects are absent, since the former is represented by a large set of model instantiations while the latter has only a single advocate (that is, the GCM).

This conundrum can be escaped by using a uniform prior on z. Attributing equal prior probability to both values of z reflects an equal belief

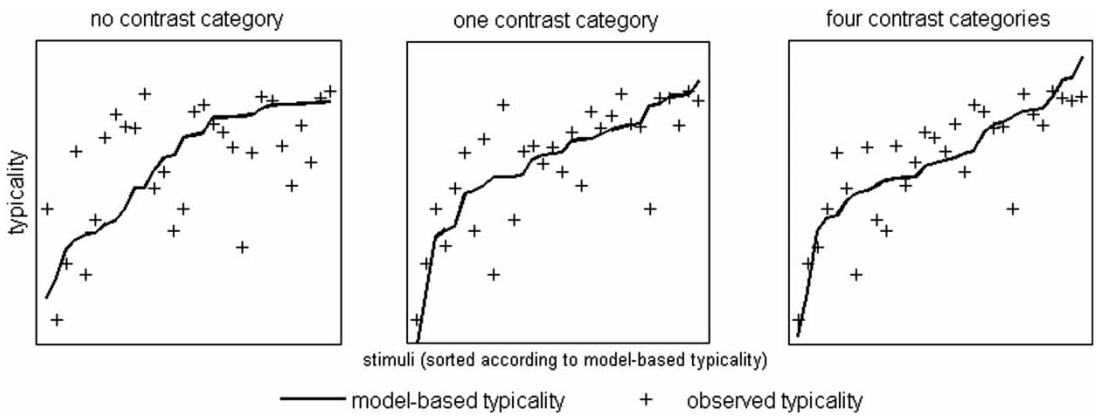


Figure 2. *Illustration of model predictions with different hyperparameter values. On the* x*-axis, all members of* mammals *are represented, rank ordered according to the model-based typicality score, and the* y*-axis represents typicality. The solid line reflects the model-based score for a particular stimulus, and the cross shows the corresponding observed typicality score. For the leftmost graph, no contrast categories are added ($z = 0$, κ and λ are irrelevant). For the middle graph,* birds *is the contrast category ($z = 1$, $\kappa = 1$, and λ selects birds). For the right graph, all contrast categories are included ($z = 1$, $\kappa = 4$, and λ selects* birds, fish, insects, *and* reptiles*). Note that the rank ordering of stimuli on the* x*-axis differs across the three graphs, as it depends on the model-based typicality scores.*

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

in the two hypotheses: The hypothesis that no contrast category is involved ($z = 0$), and the hypothesis that there is a contrast effect ($z = 1$). When comparing the two hypotheses, the posterior probability that $z = 1$ is a weighted average across all possible instantiations of the SD-GCM. In this way, the prior belief in the hypothesis that contrast effects play a role is distributed across all the different model instantiations of the SD-GCM, and differences in complexity between GCM and SD-GCM are accommodated.

Finally, the parameters that are relevant in the GCM-process—the dimensional weights w and the sensitivity parameter c—are given standard uninformative priors (see Appendix for more information).

Inference

Once behavioural data are observed, Bayesian inference updates the prior probability distributions over the parameters to their posteriors, which convey all available information regarding the parameters. These inferences are, at least conceptually, straightforward. Upon observing data, inference involves the reversal of the generative process shown in Figure 1. Rather than moving "downstream" (i.e., from the top of the scheme—parameters—in Figure 1 to the bottom—data), the goal is to move "upstream" (from data to parameters) and infer which parameter values most likely have generated the observed data. The end result is the posterior distributions on the hyperparameters. In this way, the hierarchical Bayesian framework allows inferences to easily interpretable parameters that map perfectly on the questions that were asked: whether contrast information is used (z), how many contrast categories are used (κ), and which of the potential contrast categories are used (λ). A formal description of the framework and more details on the prior distributions and calculation of the posterior distributions can be found in the Appendix.

Results

To answer the question of whether a category's internal membership structure is influenced by contrast categories, we are mainly interested in the posterior distribution of the hyperparameters, z, κ, and λ, in the hierarchical framework. From these parameters we can learn whether it is justified to add contrast categories in the account of typicality, how many contrast categories a target category most likely has, and which contrast categories are likely candidates. We address these questions on the basis of the hierarchical Bayesian analyses in the two semantic domains: animals and artefacts. The analyses were performed for each target category separately. Moreover, the semantic domains were considered separately—that is, we did not evaluate contrast effects between domains: For a particular animal category, we only allowed other animal categories as potential contrast categories, and for the artefact categories, we only considered the other artefact categories.

Evidence for contrast effects

The first question is whether the data present evidence for the existence of contrast effects. Inference to the z parameter addresses this issue, regardless of any particular contrast set. For each category, the z compares the evidence in favour of the GCM with the evidence of all instantiations of the SD-GCM. In other words, the z parameter represents the hypotheses that there is no contrast category (when $z = 0$) and that there is a contrast effect (when $z = 1$). A priori, we assumed that both models have an equal probability. Through consideration of the data, we can update the distribution of z. The resulting posterior distribution informs us about the evidence in favour of a contrast effect for the target category under consideration.

Inferences regarding the z-parameter are presented in Figure 3. For each category we present the posterior probability of the hypotheses represented by z. For 9 out of 11 categories, we find evidence in favour of contrast effects. Only for the category *weapons* is posterior probability slightly in favour of the hypothesis that no contrast effect exists. For the category *fish*, results are inconclusive.

A principled way to evaluate whether the evidence in favour of a particular hypothesis as

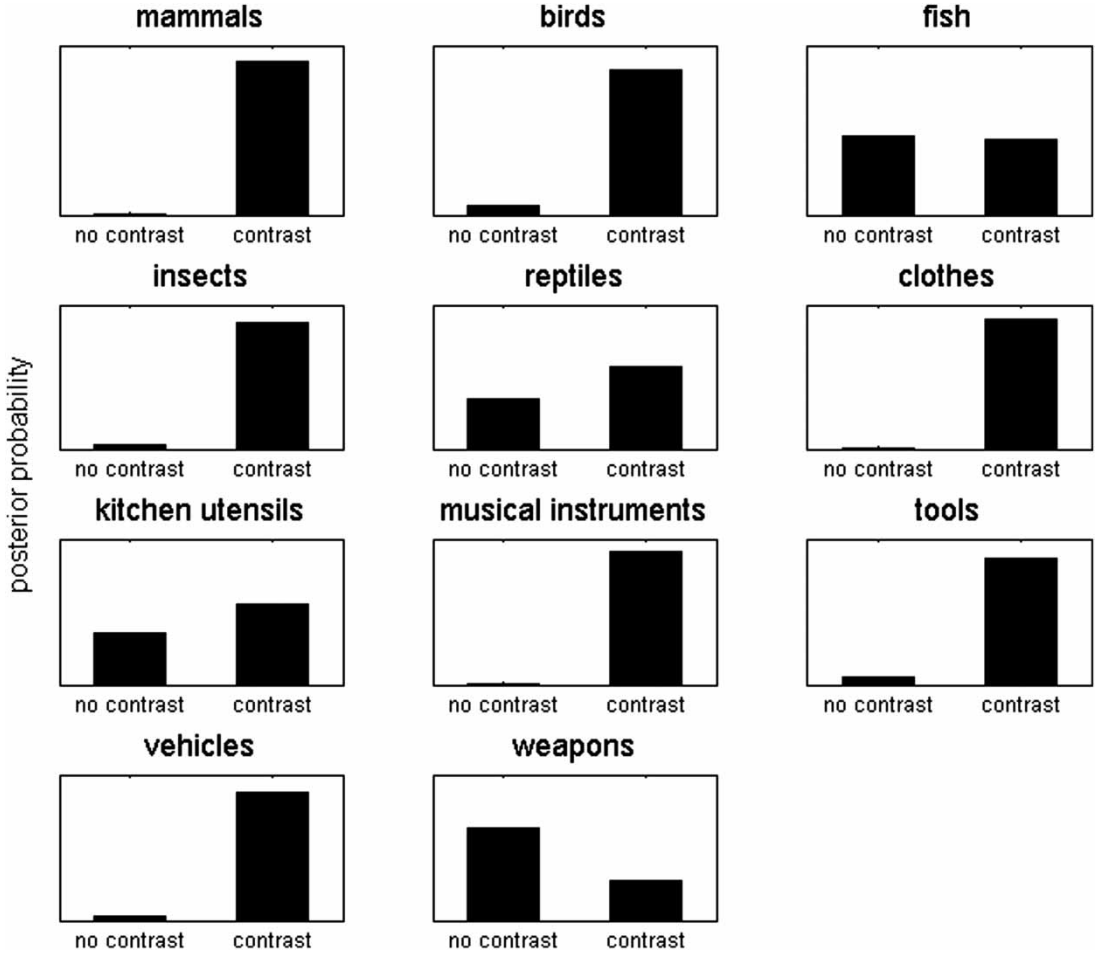


Figure 3. *Posterior probability of the hypotheses represented by* z.

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

compared to an alternative hypothesis is sufficiently large is through the Bayes factor (Kass & Raftery, 1995), which quantifies the posterior odds in favour of one hypothesis over an alternative hypothesis. Following Kass and Raftery (1995), a Bayes factor larger than 3 constitutes substantive evidence in favour of the hypothesis, reflecting that the hypothesis is more than 3 times as likely as the alternative hypothesis. Considering the Bayes factors for each category, we find sufficient evidence for contrast effects in three animal categories (*mammals*, *birds*, and *insects*) and four artefact categories (*clothes*, *musical instruments*, *tools*, and *vehicles*). For the category *weapons*, evidence was not sufficient to decide in favour of the hypothesis that there is no contrast effect.

In sum, for 7 out of 11 categories we find considerable evidence in favour of contrast effects. For the remaining 4 categories, evidence was not strong enough in favour of either hypothesis. In other words, the data are not sufficiently informative to decide for either hypothesis. Therefore, in the following section, inferences regarding the particular constitution of the contrast sets are only made for seven categories.

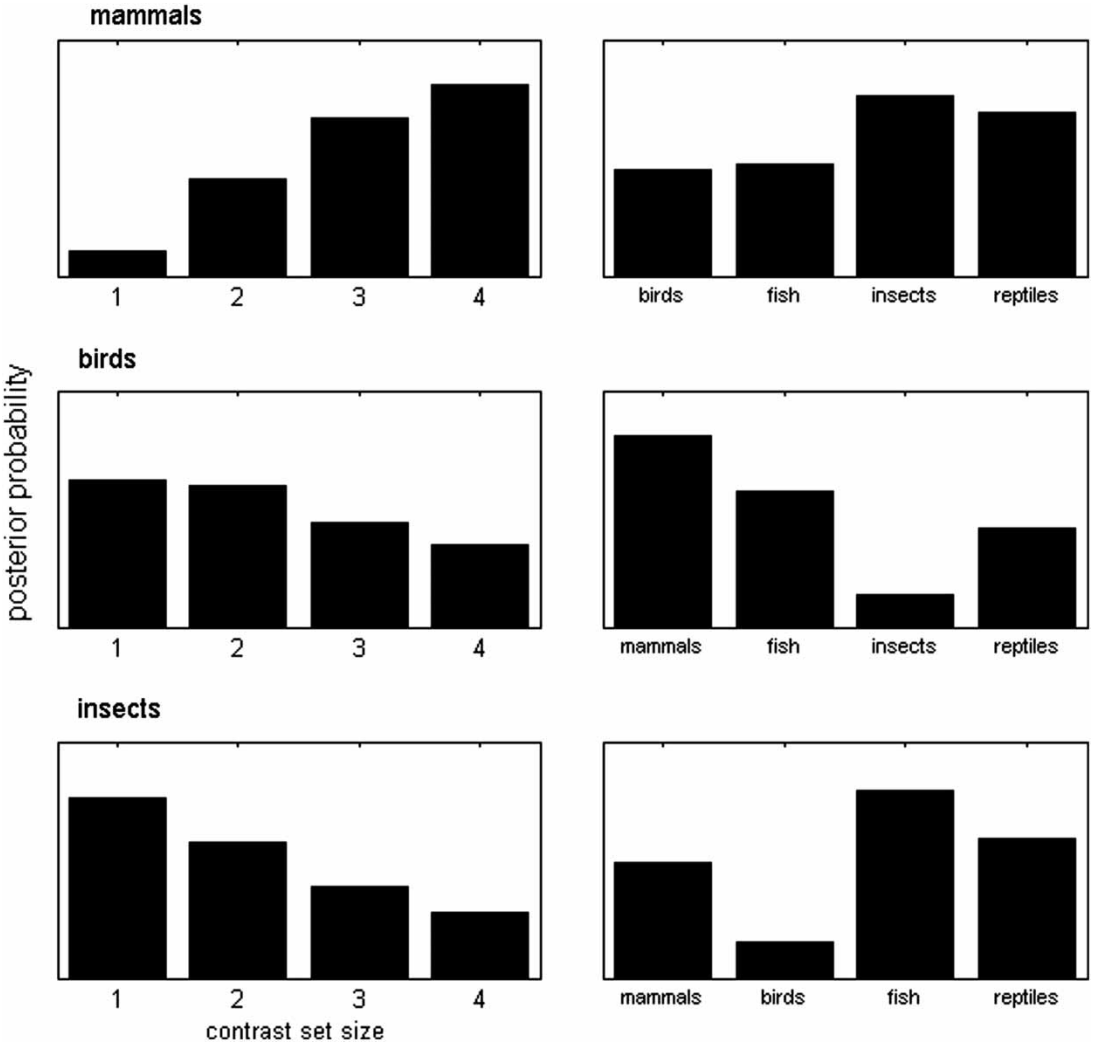


Figure 4. *Posterior probability for the κ-parameter and the λ-parameter, determining the size and the particular constitution of the contrast sets, for the three remaining animal categories.*

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

Inferences regarding the contrast set

The categories for which evidence of contrast effects was found can now be further analysed, by considering the posterior probability distributions of the two parameters that characterize the contrast set. Results are presented in Figure 4 and Figure 5. The general pattern in the posteriors reveals that a small contrast set of one or two contrast categories is most likely for the majority of remaining categories. Only *mammals* and *musical instruments* seem to evoke a larger number of contrast categories, although the evidence in the case of *musical instruments* is mixed. The results are now discussed in more detail.

Figure 4 presents the posteriors on the relevant parameters for the animal categories. The left panels show, for each category, the posterior probability of the size of the contrast set. The animal domain counted five categories. Given a particular target category, the size of the contrast set can

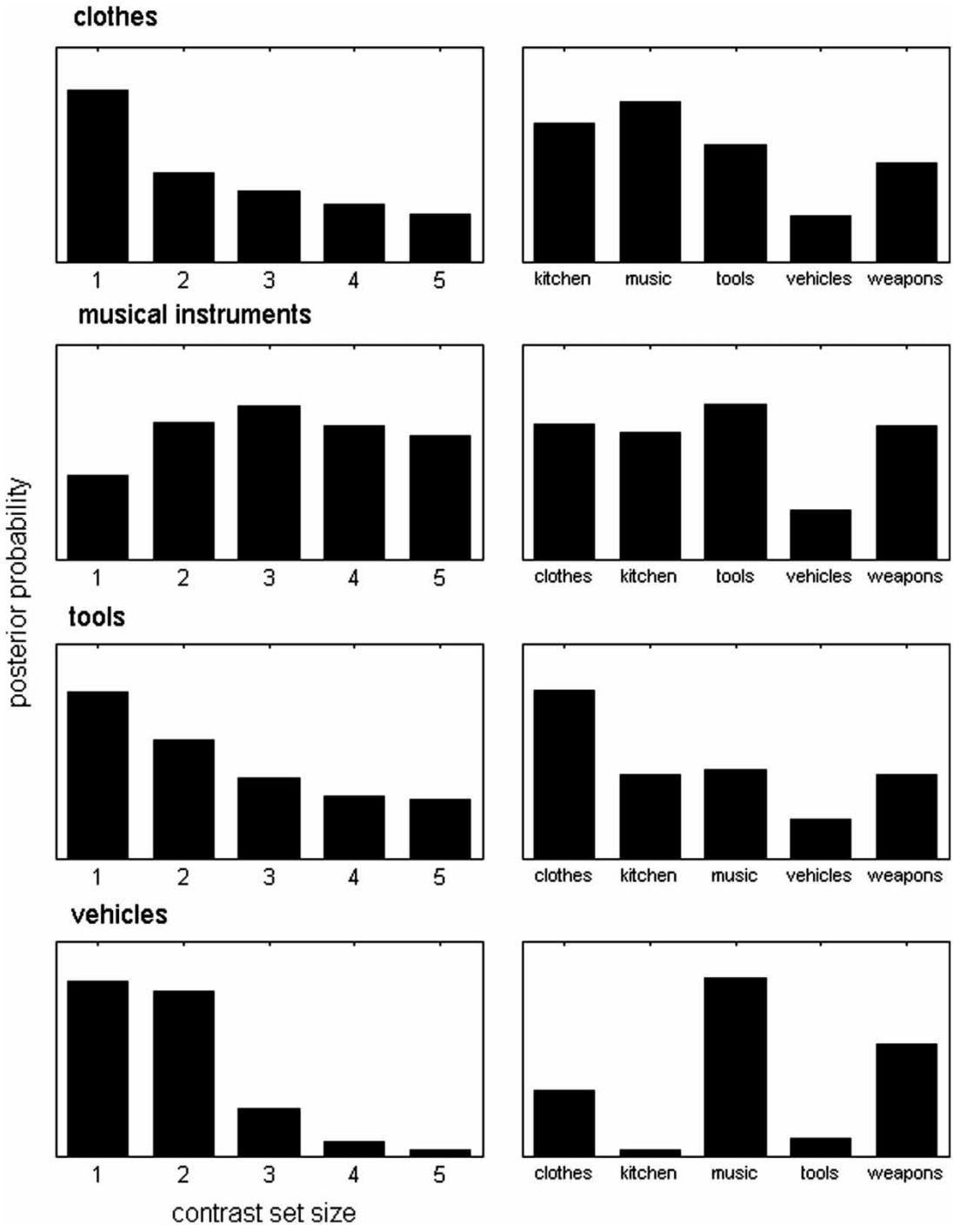


Figure 5. *Posterior probability for the κ-parameter and the λ-parameter for the four remaining artefact categories.*

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

vary from one to four. The right panels present, for each target category, the posterior probability that a particular category is in its contrast set.

For *mammals*, we can infer that larger contrast sets are most likely (with the most probable contrast set consisting of all four other categories in the domain). For *birds*, a smaller contrast set of one or two contrast categories is most likely, yet the posterior probabilities are not equally conclusive.[2] The contrast set most likely consists of the categories *mammals* and *fish*. For *insects*, again a small contrast set of one or two categories is most likely, consisting of *fish* and *reptiles* or *mammals*.

Results for the four categories in artefact domain are presented analogously in Figure 5. For *clothes*, we find evidence for a small contrast set, most likely consisting of musical instruments, *kitchen utensils*, or tools. For *musical instruments*, the κ-parameter tops at three contrast categories, but little can be inferred regarding which categories are most likely in the contrast set, except that *vehicles* is most likely not in the set. For *tools*, a small contrast set is most likely, with *clothes* being the most probable candidate for contrast category. For *vehicles*, again a small contrast set (of size one or two) is most likely, most likely consisting of *musical instruments* and *vehicles*.

Goodness of fit

Until now, different model instantiations were compared within the hierarchical Bayesian framework in terms of parameter posterior probability distributions, without considering their absolute goodness-of-fit of the model instantiations to the observed typicality judgements. In this section, we present analyses regarding the goodness of fit of the model instantiations selected on the basis of the hyperparameters presented in the previous sections. Table 1 presents for each category the correlation between observed and model-based typicality scores. In general the selected models perform well in their account of the observed typicality gradient, achieving a correlation of .74 averaged across categories. For five categories, the correlations are excellent ($r > .80$), and for only one category (*musical instruments*), the correlation drops slightly below .6.[3]

Table 1. *Correlations between observed and model-based typicality scores following from the posterior probability of the three hyperparameters* z, κ, *and* λ

Target category	*Contrast set*	r
mammals	birds, fish, reptiles, insects	.83
birds	mammals	.88
fish	—	.81
insects	fish	.74
reptiles	—	.67
clothes	musical instruments	.83
kitchen utensils	—	.61
musical instruments	clothes, tools, weapons	.59
tools	clothes	.70
vehicles	musical instruments	.82
weapons	—	.63

Note: — = no contrast category used for the model-based account.

In sum, the analyses have revealed that for most of the categories under consideration, there is clear evidence in favour of contrast effects. For most categories, we find that a small contrast set of one or two contrast categories is most likely, but the data were not always sufficiently informative to draw strong inferences (e.g., *birds*). For two categories (*mammals* and *musical instruments*), we find evidence for larger contrast sets. All in all, evidence is clearly in favour of the hypothesis that the typicality gradient of natural language categories is influenced by contrast categories, and that consideration of more than one contrast category is not implausible.

GENERAL DISCUSSION

In his seminal analysis of meaning and reference of signs in language, de Saussure (1916/1972, p. 166) states that "in language, there are only differences,

[2] Note, however, that the previous analyses have revealed the presence of contrast effects. The data, however, are less informative as to the constitution of the contrast set.

[3] Closer inspection of the scatter plots of predicted versus observed typicality scores for each category, and the corresponding residuals, did not reveal any systematic prediction errors.

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

and no positive terms. . . . In a sign, what matters more than any idea or sound associated with it, is what signs surround it".[4] While de Saussure's position may be rather extreme, the issue of the embeddedness of everyday concepts in conceptual networks does suggest the possibility that certain aspects of the concept's representation are under influence of neighbouring concepts. In the present study, we examined the graded membership in the extension of a concept.

Following Rosch and Mervis (1975), we hypothesized that the degree to which a member of a category is typical of the category is determined not only by the similarity of the member to fellow category members, but also by the dissimilarity to members of other categories—that is, we expected typicality in concepts to be under influence of contrast effects originating from neighbouring concepts. To investigate this issue, we compared two computational exemplar models in their account of the typicality gradient in 11 everyday concepts from two semantic domains, animals and artefacts. According to the GCM, typicality of a category member derives from similarity to category members. According to the SD-GCM, both similarity to category members and dissimilarity to members of alternative categories determine the typicality gradient in a category.

The model comparison and evaluation in a hierarchical Bayesian framework revealed strong evidence that contrast effects exist. For 9 out of 11 categories, evidence was in favour of contrast effects, and for 7 categories the evidence was overwhelmingly strong. For these categories, not only similarity to category members determined the typicality of a category member, but also the dissimilarity to the members of some set of contrast categories. For only one category, evidence was in favour of the absence of contrast effects, yet the evidence was not sufficiently strong to decisively draw conclusions. For most concepts that showed contrast effects, we found evidence for the influence of one or two contrast categories on their internal membership structure.

It has been argued that contrast effects are harder to demonstrate at a superordinate level than at a more basic level (Markman & Wisniewski, 1997), due to a decrease in common ground between categories when rising in the hierarchy. The lack of commonalities—that is, *alignable* differences between superordinates—would restrict the relevance of one category for another. For the same reason, it has been suggested that artefact categories are more prone to contrast effects (Goldstone, 1996): Artefact categories have more overlap, making them more mutually relevant. Our focus on superordinate categories, and the inclusion of both the artefact and animal domain, thus presented a strong test for contrast effects. Through establishing contrast effects for superordinate categories, and in finding no notable differences between the two domains, we have supplied strong support for contrast effects. The fact that contrast effects can be found even under those conditions that are considered to be hostile to contrast effect makes an impressive case for the existence for contrast effects and suggests that these effects might be more widespread than previously thought (see also Verheyen, De Deyne, Dry, & Storms, 2011).

A notable strength of the present approach to studying contrast effects is that we did not a priori choose contrast categories, but inferred which categories are contrastive for a particular target category. The inference is based on psychologically plausible models of representation and cognitive processes that have been shown to underlie typicality judgements and categorization decisions (e.g., Nosofsky, 1992; Stewart & Morin, 2007; Voorspoels et al., 2008). Having established contrast effects, a natural follow-up question is *why* certain categories are contrastive for a particular target category: Which characteristics distinguish contrast categories from noncontrastive categories? This question is particularly difficult to answer in an ecologically valid and highly complex domain such as natural language concepts, which is the focus of the present study.

[4] In the original: "mais dans la langue, il n'y a que des differences sans terms positifs. . . . Ce qu'il y a d'idée ou de matière phonique dans un signe importe moins que ce qu'il y a autour de lui dans les autres signes."

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

There are many candidate characteristics—such as feature overlap, alignability of differences, association strength between concepts, and so on—that require careful examination, but this falls outside the scope of the present study.

Implications for models of concepts

The present findings underline the necessity of considering concepts in their semantic network, at least for some tasks such as typicality ratings. Obviously, the embeddedness in such a network has consequences for each of its nodes, and this is not without consequences for the internal characteristics of the concepts. In a broader perspective, our findings highlight caution that needs to be undertaken when transposing models that are successfully tested in artificial category learning tasks to everyday concepts. In the former field a number of excellent models have been developed that carefully implement different views on category representation, but blind application in a more naturalistic context, without careful consideration of the specific characteristics of everyday concepts, might lead to incomplete and misleading conclusions. For example, the GCM is one of the most influential and most successful categorization models (see Nosofsky, 1992; Vanpaemel & Storms, 2010, for overviews) but in its account of typicality (e.g., Nosofsky, 1988) it does not (automatically) take into account the richness of the semantic domain in which a natural language category is embedded. Models that explicitly allow for contrast effect to take place, such as the SD-GCM, seem more appropriate to apply to everyday concepts/natural language categories.

Everyday concepts are indeed quite different from artificial categories that are learned in the lab (e.g., Malt & Smith, 1984; Murphy & Medin, 1985). Tasks that involve everyday concepts generally require "a whole bunch more cognition" than the "tiny sliver of cognition" that a computational model generally focuses on (Murphy, 2010, p. 302) and that can be zoomed in on in categorization experiments. By consequence, the models that are used in concept research have to take into account this increased complexity. One way to do this is through the use of hierarchical methods (Lee, 2011). The present analyses highlight that extra levels of complexity can be implemented and added to relatively simple computational models that are thought to underlie the phenomenon at hand, thus providing a powerful tool to tackle more complex domains of reality. By presenting an example of the application of these methods, this research adds to bridging the existing gap between the field of artificial categorization with its well-defined models and the complexity of the field of everyday concepts.

Original manuscript received 05 October 2011
Accepted revision received 23 December 2011
First published online 27 April 2012

REFERENCES

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

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicists' challenge in cognitive science. *Cognitive Science*, *22*, 295–318.

Block, N. (1999). Functional role semantics. In R. A. Wilson, & F. C. Keil (Eds.), *MIT encyclopedia of the cognitive sciences* (pp. 331–332). Cambridge, MA: MIT Press.

Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, *8,* 240–247.

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

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

de Saussure, F. (1972). *Cours de linguistique général* [Course in general linguistics]. Paris, France: Edition Payot (Original work published 1916).

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

Dietrich, E., & Markman, A. B. (2003). Discrete thought: Why cognition must use discrete representations. *Mind & Language*, *18,* 95–119.

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

Goldstone, R., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31(2),* 169–180.

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

Hampton J. (1993). Prototype models of concept representation. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67–95). San Diego, CA: Academic Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90,* 773–795.

Keller, D., & Kellas, G. (1978). Typicality as a dimension of encoding. *Journal of Experimental Psychology: Human Learning and Memory*, *4,* 78–85.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychonomic Bulletin*, *112,* 500–526.

Kruskal, J., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104,* 211–240.

Lee, M. (2004). A Bayesian analysis of retention functions. *Journal of Mathematical Psychology*, *48,* 310–321.

Lee, M. D. (2011). How cognitive science can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55,* 1–7.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning & Verbal Behavior*, *23,* 250–269.

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

McNamara, T. P., & Miller, D. L. (1989). Attributes of theories of meaning. *Psychological Bulletin*, *106,* 355–376.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Millikan, R. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT.

Murphy, G. L. (2010). The contribution (and drawbacks) of models to the study of concepts. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 299–312). Cambridge, UK: Cambridge University Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92,* 289–316.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44,* 190–204.

Nosofsky, R. M. (1984). Choice, similarity and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *10,* 104–114.

Nosofsky, R. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115,* 39–57.

Nosofsky, R. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *14,* 54–65.

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

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

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

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, *14,* 665–681.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104,* 192–233.

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

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237,* 1317–1323.

Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, *49,* 403–409.

Stewart, N., & Morin, C. (2007). Dissimilarity is used as evidence of category membership in multidimensional

perceptual categorization: A test of the similarity–dissimilarity generalized context model. *The Quarterly Journal of Experimental Psychology*, *60,* 1337–1346.

Vanpaemel, W. (2011). Constructing informative model priors using hierarchical methods. *Journal of Mathematical Psychology*, *55,* 106–117.

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

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

Verheyen, S., De Deyne, S., Dry, M. J., & Storms, G. (2011). Uncovering contrast categories in categorization with a probabilistic threshold model. *Journal of Experimental Psychology: Learning, Memory & Cognition, 37,* 1515–1531.

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

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

APPENDIX

A hierarchical Bayesian approach

We give an overview of the hierarchical Bayesian framework applied in our analyses in search of contrast effects. In short, the hierarchical Bayesian framework is used to analyse a large family of models of typicality for a particular target category, where each model is defined by a particular set of contrast categories, the contrast set. The framework allows inference to parameters that determine whether a contrast set is used in the typicality calculation and how the contrast set is constituted (that is, how many categories and which categories make up the contrast set).

Figure A1 provides a graphical model of the hierarchical Bayesian framework applied in our analyses. In a graphical model, the nodes represent variables, and the graph structure is used to indicate dependencies between variables, children nodes depending on parent nodes. Following general conventions, continuous variables are represented by circular nodes and discrete variables by square nodes. Observed variables are indicated by shading of the node. Stochastic and deterministic unobserved variables are distinguished by using single and double borders, respectively. We also use plate notation, enclosing a subset of the graph with square boundaries that have independent replications in the framework.

The graphical model has three levels. The top level displays the geometric stimulus representation and the mechanism that generates a particular contrast set. The middle level displays the mechanism of the GCM and SD-GCM, involving the derivation of similarity from distance in the stimulus representation. The lower level displays the observed typicality ratings. For each level, we discuss the relations between the nodes and the prior probability distributions on the parameters involved.

The top level: Constructing a contrast set

At the top of the graphical model sits the geometric stimulus representation S, containing the coordinates of all exemplars of a particular domain, the particular combination of contrast categories C, and the two hyperparameters, κ and λ, which together determine the composition of C. The κ parameter determines the number of categories in the contrast set and is an integer running from 1 to the maximum of categories in the contrast set. The λ parameter determines the probability that a particular category is in the contrast set, for each category taking values from 0 to 1 and summing to 1 across categories. For a particular value of κ and λ, we can derive a probability for each of the possible contrast sets.

While the geometric stimulus representation is an observed entity, we are ignorant as to the precise values of the κ and λ parameter. We use the following priors on these parameters:

$$\kappa \sim DiscreteUniform\left(\frac{1}{N_c}\right)$$

$$\lambda \sim Categorical\left(\frac{1}{N_c}, \frac{1}{N_c}, \cdots \frac{1}{N_c}\right),$$

where Nc represents the number of potential contrast categories. These priors express the a priori ignorance regarding how many categories are in the set (each number, ranging from one to the maximum number of categories, is equally likely) and which categories are in the set (each category is equally likely to be in the set).

The middle and bottom levels: Calculating typicality from similarity

The middle level of the graphical model displays the derivation of similarity from distance in the stimulus representation. The calculation of similarity between two exemplars depends on their weighted distance in the stimulus representation (for both the GCM and SD-GCM):

$$S_{ij} = \exp(-cd_{ij})$$

where c is the sensitivity parameter. The weighted distance between two exemplars in turn is derived from the coordinates

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012

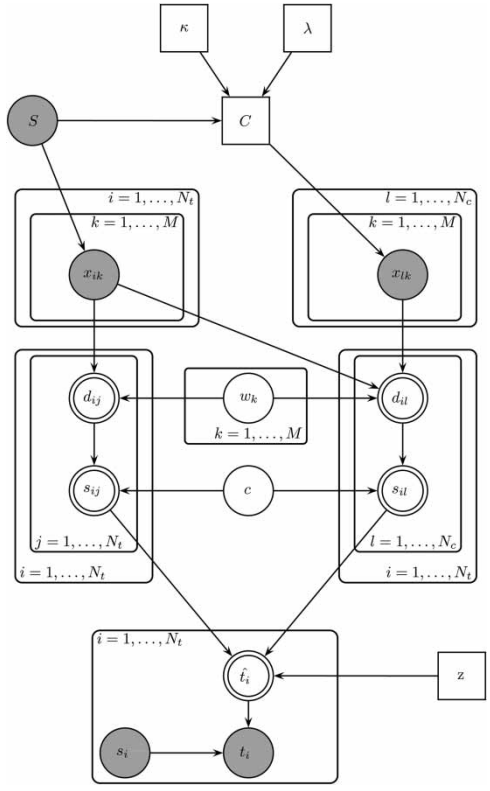


Figure A1. *Graphical model of the hierarchical Bayesian framework applied in our analyses.*

of the exemplars:

$$d_{ij} = \left(\sum_{k=1}^{M} w_k |x_{ik} - x_{jk}|^2 \right)^{1/2}$$

where x_{ik} and x_{jk} are the coordinates of exemplars i and j on dimension k, w is a parameter reflecting the attention weight for dimension k, and M is the number of dimensions. The parameters that influence the similarity in both the GCM and SD-GCM are the dimension weights w and the sensitivity parameter c. We used standard noninformative priors on these parameters—that is, for the weights a uniform (0, 1) distribution, with the restriction that the weights sum to 1, and c is gamma (.001, .001) distributed.

The model-based typicality scores, $\hat{t}_{iA}$ rely either on the standard GCM, not using the contrast set, or on the SD-GCM, relying on similarity to the target category and dissimilarity to the contrast set. Formally, typicality of an exemplar i for a category A is the summed similarity to all target category members according to the GCM:

$$\hat{t}_{iA} = \sum_{j \in A}^{N_A} S_{ij}$$

where j is a member of the target category A, and N_A is the number of exemplars in category A. In the graphical model, the similarity to members of the target category is represented in the left track at the middle level. The right track reflects the contrast component—that is, the dissimilarity to members of the contrast set—which is added to the calculation of typicality according to the SD-GCM:

$$\hat{t}_{iA} = \sum_{j \in A}^{N_A} S_{ij} + \sum_{j \in C}^{N_c} (1 - S_{il})$$

where N_c is the number of exemplars in the contrast set C.

The parameter z determines whether the typicality scores of a target category depend on only similarity to target category exemplars (when z is 0) or both similarity to target category members and dissimilarity to exemplars of the contrast set. In other words, z can have values 0 and 1, deciding whether the GCM or the SD-GCM is used to calculate t_{iA}. A priori, we are ignorant as to whether the dissimilarity component is used into the typicality calculation, translating into the following prior:

$$z \sim Bern(.5)$$

At the bottom level of the graphical model, the typicality ratings can be found. We assume that each of the observed typicality ratings t_{iA} comes from a Gaussian distribution with mean $\hat{t}_{iA}$—that is, the model-based typicality score—and variance σ_i^2:

$$t_{iA} \sim N(\hat{t}_{i,A}, \sigma_i^2)$$

Where σ_i is estimated by s_i—that is, the standard deviation across participants of the typicality judgements for exemplar i, reflecting the data precision (see, e.g., Lee, 2004).

Computing the posteriors

Ultimately, we are interested in inferences to three parameters in the hierarchical framework: z, κ, and λ. These parameters are informative as to the existence of contrast effects (z) and the composition of the contrast set: how many (κ) and which (λ) categories are in the set. The parameter posteriors are computed

using Bayes' rule:

$$p(z|t) = \frac{p(t|z)p(z)}{p(t)},$$
$$p(\kappa|t) = \frac{p(t|\kappa)p(\kappa)}{p(t)},$$
$$p(\lambda|t) = \frac{p(t|\lambda)p(\lambda)}{p(t)},$$

with t referring to the observed typicality scores. The posterior probability of the parameters thus relies on the prior probability of a parameter value and the probability that the typicality scores are observed given the parameters' values. The latter probabilities can be calculated by summing across all possible models, weighted by the models' probability given the parameter at hand:

$$p(t|z) = \sum_j p(t|M_j)p(M_j|z),$$
$$p(t|\kappa) = \sum_j p(t|M_j)p(M_j|\kappa),$$
$$p(t|\lambda) = \sum_j p(t|M_j)p(M_j|\lambda),$$

where the probability of a particular model can be derived from the parameter priors of z, κ, and λ. The marginal likelihood of a particular model $p(t_j \mid M_j)$ considers all possible parameter values w_k and c. This is done by integrating the probability of the data across all possible parameter combinations, weighted by the prior probability of the parameter combinations $p(w, c)$:

$$p(t|M) = \int p(t|M, w, c)p(w, c)\mathrm{d}w\mathrm{d}c$$

On the basis of the Gaussian distribution of the typicality scores, a probabilistic measure for a particular model M_j—that is, $p(t_j \mid M_j)$—and a particular set of parameter values w and c is given by:

$$p(t|M, w, c) = \prod_i^{N_A} \frac{1}{S_i\sqrt{2\pi}} \exp\left(-\frac{(t - \hat{t}_i)^2}{2S_i^2}\right)$$

To calculate the term, we transformed both model-based scores ($\hat{}t_i$) and observed scores (t_i) to the same scale through standard normalization. We used a straightforward numerical method to approximate the integral that defines the marginal likelihood. By sampling a large number of parameter values according to their prior distribution, the marginal likelihood can be estimated by:

$$p(t|M) \approx \frac{1}{R}\sum_{r=1}^{R} p(t|M, w_r, c_r)$$

The sampling was done using a simple Monte Carlo method. First, parameter values were independently sampled according to their prior distribution. For each consecutive sample, the model likelihood was derived, and the average across all previous likelihoods (that is, all previous samples and the current sample) was calculated. After a large number of samples (we used 40,000 samples in a chain), this average converges to the marginal likelihood. Convergence was evaluated by visual inspection. For each estimate, two independent chains were sampled. The resulting estimates were in all cases nearly identical.

Downloaded by [K.U.Leuven - Tijdschriften] at 08:57 30 April 2012