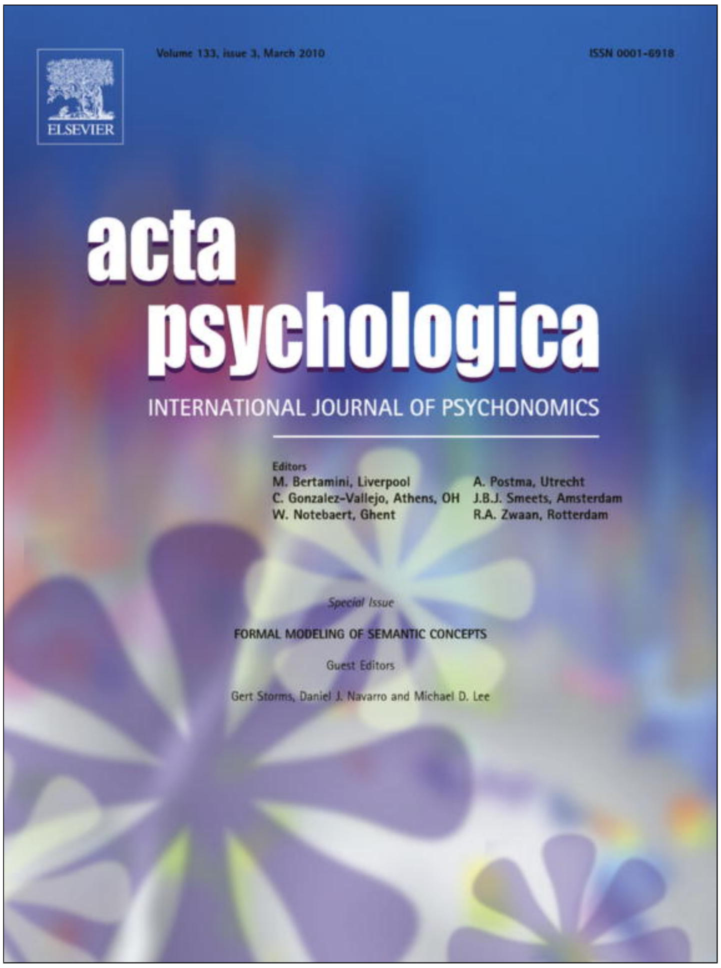
Volume 133, issue 3, March 2010
ISSN 0001-6918
ELSEVIER
acta
psychologica
INTERNATIONAL JOURNAL OF PSYCHONOMICS
Editors
M. Bertamini, Liverpool
C. Gonzalez-Vallejo, Athens, OH
W. Notebaert, Ghent
A. Postma, Utrecht
J.B.J. Smeets, Amsterdam
R.A. Zwaan, Rotterdam
Special Issue
FORMAL MODELING OF SEMANTIC CONCEPTS
Guest Editors
Gert Storms, Daniel J. Navarro and Michael D. Lee

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Finding the features that represent stimuli

Matthew D. Zeigenfuse [*], Michael D. Lee

Department of Cognitive Sciences, University of California, Irvine, CA 92697-5100, USA

ARTICLE INFO

Article history:
Received 19 December 2008
Received in revised form 22 May 2009
Accepted 19 July 2009
Available online 11 September 2009

Keywords:
Similarity
Mental representation
Feature representation
Categorization
Bayesian modeling

ABSTRACT

We develop a model for finding the features that represent a set of stimuli, and apply it to the Leuven Concept Database. The model combines the feature generation and similarity judgment task data, inferring whether each of the generated features is important for explaining the patterns of similarity between stimuli. Across four datasets, we show that features range from being very important to very unimportant, and that a small subset of important features is adequate to describe the similarities. We also show that the features inferred to be more important are intuitively reasonable, and present analyses showing that important features tend to focus on narrow sets of stimuli, providing information about the category structures that organize the stimuli into groups.

1. Introduction

A common assumption in cognitive psychology is that stimuli can be represented in terms of the presence or absence of a set of features. This assumption provides a representational foundation for many models of higher-level cognitive processes, including models of memory (e.g., Dennis & Humphreys, 2001; Hintzman, 1984; Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997), category learning (e.g., Lee & Navarro, 2002; Love, Medin, & Gureckis, 2004; Medin & Schaffer, 1978), and decision-making (e.g., Gigerenzer & Goldstein, 1996; Payne, Bettman, & Johnson, 1990). It also serves as a useful basis for understanding mental representations themselves, since it makes possible formal analyses. Having assumed that stimuli are represented by features, it is possible to develop quantitative measures of how features relate to each other, how features distribute themselves across categories, and so on.

In this paper, we investigate both of these aspects of feature representation, and so pursue related methodological and theoretical goals. Methodologically, we develop a model for finding feature representations, and apply it to the Leuven Concept Dataset (De Deyne et al., 2008). Theoretically, we explore what the features found by our model can tell us about the nature of mental representation. We show the representations found by our model are reasonable and the included features are likely to be important to participants judgments of similarity. Moreover, we explore whether the features people use tend to belong to many stimuli or few, how they relate to category structures over stimuli, and which sorts of features are the most salient ones.

The outline of the paper is as follows. First, we provide the intuition behind and formal characterization of the model for finding feature representations. Second, we demonstrate the model on a toy problem. Third, we apply the model to the Leuven Concept Database, describing the feature representations it finds, and their ability to account for the available empirical data. Finally, we analyze the properties of the feature representations found for the Leuven data, and discuss their implications for what makes a feature useful.

2. A model for finding feature representations

The model we develop is an extension of one originally proposed by Zeigenfuse and Lee (2008). Their approach involved combining two existing methods for finding feature representations, aiming to maintain the best aspects of both, and overcoming the weaknesses in each. The first existing method is the feature generation task, in which participants generate a set of candidate representational features, and then judge whether or not each stimulus has each of the features. The second existing method is the similarity judgment task, in which participants assess—via rating scales for the Leuven data, although other methods are possible—the psychological similarity between each pair of stimuli. These similarities can then be used by models like additive clustering (Shepard & Arabie, 1979), or its various extensions (e.g., Navarro & Lee, 2004) to infer a feature set, and the assignments of each feature to each stimulus.

As noted by Zeigenfuse and Lee (2008), both these tasks have strengths and weaknesses as methods for finding the features that represent stimuli. Feature generation immediately provides stimu-

[*] Corresponding author. Tel.: +1 949 824 4353; fax: +1 949 824 2307.
E-mail address: mzeigenf@uci.edu (M.D. Zeigenfuse).

0001-6918/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2009.07.014

lus representations, but relies on the ability of participants to generate and assign the features. The free-form nature of the generation task is especially problematic, because it is not clear how people generate the features, nor how they decide to stop listing candidate features. Making inferences from similarity judgments has a sounder theoretical basis, but is more computationally challenging (see Navarro and Griffiths (2004), for the state of the art). Perhaps most fundamentally, additive clustering models do not provide any semantic interpretation of the features they derive.

The key insight underlying the new method proposed by Zeigenfuse and Lee (2008) was that the observed data in feature generation and similarity judgments tasks can be related to each other, because both types of data are based on the same underlying representation. When people do feature generation, they rely (in part) on the same features they use to make similarity judgments. Through this relationship, similarity judgments provide information to decide which features, from a large generated set, are the important ones for representing stimuli. In other words, the features from a generation task that are worth keeping to represent stimuli are those features needed to explain the similarities between the stimuli. This combined approach has the dual advantages of finding just those features that are important for representing stimuli, and providing a meaningful semantic label for these features.

2.1. Formalization of the model

The model developed by Zeigenfuse and Lee (2008) formalized the idea of combining tasks by making specific assumptions about how the generation and similarity tasks work. In this paper, we rely on basically the same ideas, although we make a few natural improvements, which we will highlight in the formal description of the model. In essence, for feature generation, we assume that when people produce feature lists, they provide all of the important features in their mental representations of the stimuli, but augment these with additional peripheral features. For similarity judgment, we assume people use common features, in the way formalized by additive clustering models (Shepard & Arabie, 1979).

The common-features model is far from the only possible choice for a model of similarity judgments. For example, rather than assume people start with stimuli being completely dissimilar and accumulate similarity with common features, we could assume people start with stimuli being completely similar and decrease similarity with each distinctive feature (e.g., Medin & Schaffer, 1978; Restle, 1959). Alternatively, we could assume people use some combination of common and distinctive features, such as a Contrast model (Tversky, 1977) or ratio model (Bush & Mosteller, 1951; Eisler & Ekman, 1959; Gregson, 1975; Tenenbaum & Griffiths, 2001). Our decision to use a common-features model here is based on its simplicity and empirical success in previous modeling of similarity data.

We denote a feature representation of n stimuli using m features by a matrix $F = [f_{ik}]$, where $f_{ik} = 1$ if the ith stimulus has the kth feature, and $f_{ik} = 0$ if it does not. Using this notation, our account of feature generation is that a $n \times (m + m^+)$ matrix F^+ is generated, containing the m true features, and an additional m^+ peripheral features.

We denote the pairwise similarities between n stimuli, as provided by the pth participant, by a matrix $S_p = [s_{ijp}]$, where s_{ijp} is the similarity between the ith and jth stimuli given by the pth participant. We denote the weight of the kth feature as w_k, and the constant baseline level of similarity shared by all stimuli as c.

Finally, we introduce a set of latent indicator variables, one for each feature, whose role it is to indicate whether each feature is a true feature or an additional feature. We denote the latent indicator for the kth feature by z_k, with $z_k = 1$ if the kth feature is part of the underlying representation, but $z_k = 0$ if the kth feature is an additional feature produced during the generation task. This means that the model estimates the similarity between the ith and jth stimuli as

$$\hat{s}_{ij} = \sum_{k=1}^{(m+m^+)} z_k w_k f_{ik} f_{jk} + c \tag{1}$$

This equation is easy to interpret. It assumes a common-features model of similarity, with the weights of shared features adding to give the similarity between any pair of stimuli. Critically, though, only those features indicated to belong to the true underlying representation are involved in this similarity process. The kth feature only influences the similarity between stimuli if $z_k = 1$. Intuitively, this means that the inferences the model makes about the z_k indicator variables correspond to 'pruning', 'paring back', or 'regularizing' the whole list of generated features to a smaller set of true, useful or important features, based on the information latent in the measures of stimulus similarity.

2.2. Bayesian inference for the model

To make inferences using our model, we implemented it as a probabilistic graphical model (see Jordan (2004) and Lee (2008), for statistical and psychological introductions, respectively) in WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000). Graphical models provide a framework for formally expressing dependencies among parameters of a model. When coupled with a set of conditional distributions they unambiguously specify the likelihood of the observed data and the prior distribution over model parameters. Given data, they then specify a posterior distribution over the model parameters via Bayes rule which can be sampled using Markov Chain Monte Carlo (MCMC) methods. WinBUGS (Lunn et al., 2000) is software that generically implements MCMC sampling for graphical models.

Fig. 1 shows the graphical model, using the same notation as Lee (2008). Nodes in the graph correspond to variables, and edges indicate dependencies, with children depending on their parents. Shaded and unshaded nodes correspond to observed and unob-

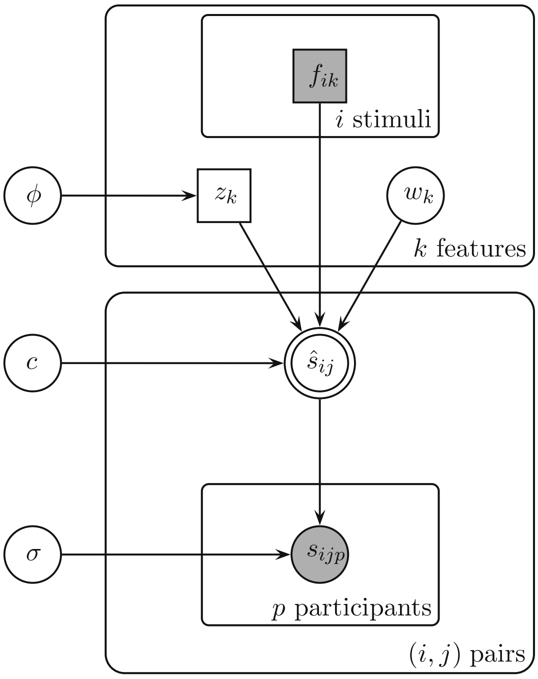

Fig. 1. Graphical representation of our model.

served variables. Circular and square nodes correspond to continuous and discrete variables. Double bordered nodes are deterministic, and are just included to aid interpretation of the model. Finally, encompassing plates are used to show independent replications of the graph structure within the model.

The two observed variables in the graphical model are the features f_{ik} from the generation task, and the similarities s_{ijp} from the similarity task. For each pair of stimuli, the known features combine with the unknown latent indicators z_k and weights w_k to give their modeled similarity $\hat{s}_{ij}$ according to Eq. (1).

We then make the standard assumption (e.g., Lee, 2002; Tenenbaum, 1996) that the individual participant observed similarity data are noisy, according to a Gaussian error model with common variance, so that

$$\hat{s}_{ijp} \sim \text{Gaussian}\left(s_{ij}, \sigma^2\right) \tag{2}$$

To specify prior distributions, we follow the advice of Gelman (2006) and place a uniform prior on the variance:

$$\sigma \sim \text{Uniform}\,(0, 10) \tag{3}$$

This allows the noisiness of the similarity judgments to be inferred from the repeated-measures provided across participants. We also place uniform priors on the feature weights.

$$W_k,\ c \sim \text{Uniform}\,(0, 1) \tag{4}$$

Finally, we place an unknown rate on the latent indicators, consistent with the idea that there is an unknown base-rate with which stimuli have features.

$$z_k \sim \text{Bernoulli}\,(\phi) \tag{5}$$

with a uniform prior on the rate itself, so that

$$\phi \sim \text{Uniform}\,(0, 1) \tag{6}$$

This graphical model extends that used by Zeigenfuse and Lee (2008) in two ways. First, using individual participant data to infer the common variance extends the original approach of using averaged similarity data and making a fixed assumption about the variance. Second, introducing the base-rate parameter ϕ extends the original approach of fixing the base-rate at 0.5.

3. A demonstration of the model

In this section we provide an illustrative application of the model. We created a 'toy' domain with a few stimuli and features, and constructed similarity data using the known features and weights. We then test the ability of the model to recover the known true features and their weights. This is essentially the same demonstration presented in Zeigenfuse and Lee (2008), but it worth revisiting, because of the extensions to the model we have introduced, and to provide an easily understood example of the application of the model.

3.1. Features and similarities

Table 1 gives a feature representation for four animals—a dog, cat, elephant, and monkey—in terms of seven features. Three of these features, "kept as a pet", "is hunted", and "can be dangerous", are given non-zero weights in the representation, and so correspond to true features that are an integral part of the representation of the animals. Three other features, "has a prime number of letters", "is a US political mascot", and "does not end in the letter t" are given zero weights, to indicate that they are additional features. These additional features might be able to be produced in a generation task, but are not an important part of how animals are represented. The final feature is the constant, with a weight that gives the level of similarity all animals share.

Table 1
A feature representation of four animals using three 'true' features, three 'additional' features, and a similarity constant.

Feature	Weight	Dog	Cat	Elephant	Donkey
Kept as pet	0.5	●	●		●
Is hunted	0.2		●	●	
Can be dangerous	0.1		●	●	●
Prime number of letters	–	●	●		
US political mascot	–			●	●
Does not end in "t"	–	●			●
Constant	0.05	●	●	●	●

Using the representation in Table 1, we generated true underlying pairwise similarities, using the common-features approach to similarity. That is, we added the weights of the common features for each pair of animals, to produce the similarity matrix.

$$\hat{S} = \begin{bmatrix} - & 0.55 & 0.05 & 0.55 \\ & - & 0.35 & 0.65 \\ & & - & 0.15 \\ & & & - \end{bmatrix} \tag{7}$$

We then generated artificial observed similarity matrices S for 9 simulated participants, by adding Gaussian noise with mean 0 and variance 0.05^2 independently to each cell in $\hat{S}$ for each participant.

3.2. Modeling results

We applied our model to these individual participant similarity matrices, and the feature matrix F^+ given by Table 1. As with all of our analyses in this paper, the results are based on four chains each with 4000 recorded samples. The recorded samples are treated as draws from the full joint posterior distribution of the weights $\boldsymbol{w} = (w_1, \ldots, w_k, c)$, indicator variables $z = (z_1, \ldots, z_k)$, the noise parameter σ and the base-rate parameter ϕ.

The key analysis of the model's inferences involves the joint posterior over the indicator variables. Of the 2^6 possible combinations that could be sampled (i.e., all possible patterns of features being true versus additional features), only four were ever sampled with non-negligible probability (i.e., with a proportion of at least 0.01). These four patterns are shown in Table 2, together with their proportion in the sample, which approximates their posterior mass. Each pattern corresponds to a different inference about which features are true and which are additional, and the mass measures the certainty with which each combination is the correct pattern.

The MAP assignments (i.e., the assignments with the most posterior mass) given by pattern 1 dominate the posterior, and have the right structure. In particular, the first three features—pet, hunted and dangerous—are assigned as true features, while the others are assigned as additional features, following the design we used to generate the data.

The posterior distribution of the weights conditional on the assignments given by pattern 1 also show the model is making the right inferences. The marginal expected values for the weights of the true features and constant were $w_1 = 0.49$, $w_2 = 0.16$, $w_3 = 0.11$, and $c = 0.05$, all of which are close to the original values in Table 1. In addition the expected value of the noise parameter σ and base-rate parameter ϕ were 0.05 and 0.53, respectively, which are consistent with the known way the data were generated.

This simple example illustrates how our model identifies just those generated features that play a role in the judgment of stimulus similarity. While it is possible to characterize animals, or any other stimuli, in terms of an endless number of features, only some features are important for representing and understanding. In this example, features like "is hunted" were inferred to be important in explaining similarity, while features like "has a prime number of letters" were inferred to be unimportant.

Table 2
The four patterns of indicator variable assignments with non-negligible mass found in the posterior samples, together with their proportion in the sample.

Pattern	Proportion	Pet	Hunted	Dangerous	Prime	Mascot	Not end "t"
1	0.79	•	•	•			
2	0.15	•	•	•	•		
3	0.03	•	•	•		•	
4	0.01	•	•	•			•

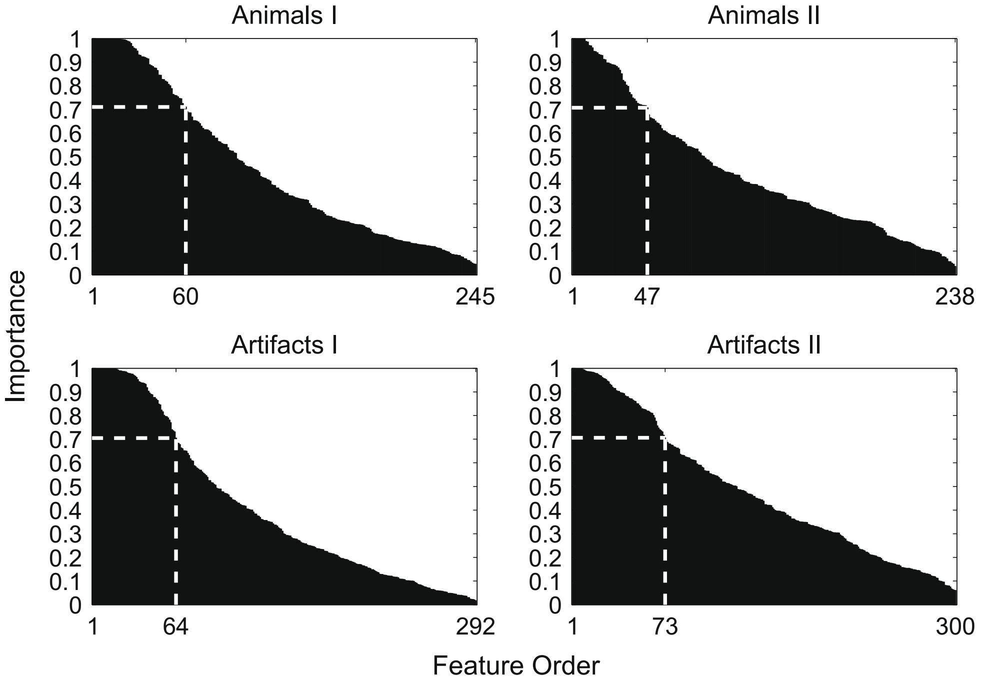

Fig. 2. The distribution of importance for the four datasets.

4. Model results for the Leuven Concept Datasets

We now turn to applying our model to the Leuven Concept Datasets. This is straightforward, because the same information used in the toy example—the feature matrix assigning features to stimuli, and the individual participant similarities between pairs of stimuli—are directly available.

Of the many available sets of stimuli for which similarity data are available, we focused on the sets using stimuli that spanned categories. That is, unlike Zeigenfuse and Lee (2008), we did not use similarity data involving all the stimuli in a category (i.e., all of the mammals), but rather on similarity data involving a selection of stimuli from each category (e.g., some mammals, some fish, some insects, and so on). We focused on these datasets because wanted to consider how important versus unimportant features related to category structures, and so needed stimuli that spanned a range of different categories.

There are four Leuven datasets of this type: two from the animal domain and two from the artifact domain. All of the datasets have five randomly chosen stimuli from each category within the domain. This means there were 25 stimuli in both of the animal datasets, and 30 in the artifact ones. Because only a subset of all the animal or artificial stimuli are involved in each individual dataset, there are a number of features that are not distinguished. That is, there are some features that have identical patterns of belonging to the stimuli. These features were only included once in the observed data for the modeling, and were given the label we thought was most meaningful (e.g., we choose "has hoofs" rather than "stands in the stable"). Note also that whether or not a feature is distinguished depends on exactly which stimuli are involved, and so there are some features that are indistinguishable in one dataset, but not another. For example, the features "is a vehicle" and "has seats" are indistinguishable in the first artifact dataset, but not in the second, because of the presence of the stimulus "wheelbarrow" in the second dataset.

4.1. Distribution of importance

A first basic question from our modeling is to ask whether there is evidence that some features are more important than others. Our model is premised on the idea that, in a feature generation task, many more features are produced than are needed to represent the stimuli in the basic cognitive context afforded by similarity comparison.

We chose to measure the importance of a feature by how often it was included to represent the stimuli in a dataset, independent of any other feature. Formally, this corresponds to the marginal posterior mass of the z_k indicator variable for the kth feature, which is approximated by the mean of z_k in the posterior sample.[1]

[1] Ideally, of course, we would be able to assess combinations of features (i.e., consider their dependencies, and not just consider them as independent) in the posterior sample. However, our experience was that the very large space of possibilities—the 2^{292} for the first animal dataset is about 8×10^{87}, for example—does not have a concentrated region of posterior mass. Rather, it seems that large regions, corresponding to combinatorial variations of the presence or absence of individual features around a good subset of features, all have some posterior mass. This sort of posterior structure seems best summarized by considering the mass of each feature independently, and makes for a much more computationally (and presentationally) tractable analysis.

Fig. 2 shows the distribution of importance for all of the features—ordered from most to least important—in all four datasets. It is clear that, for all datasets, some features were always part of the feature representation, while others rarely were, and so there is considerable variation in the importance of features. In general, the importance of features falls away fairly rapidly.

Although there is no unambiguously clear 'elbow' in the distributions, we chose the point of 0.7 importance as being a reasonable cut-off point to give an operational definition to the idea of 'important' or 'useful' features. This cut-off point is shown by the broken white lines for each dataset, and corresponds to the most important 60, 47, 64 and 73 features our of 245, 238, 292 and 300 total features. In other words, about a quarter of all features at most are included as part of stimulus representation more than 70% of the time. We use this arbitrary but reasonable cutoff in several of our later analyses, to the illustrate the idea that just a subset of the most important features can usefully form the representational basis for the stimulus sets.

4.2. Fitting similarity

Fig. 3 shows the fit of our model to the empirical similarity data for all four datasets. Both the fit using all of the available features, and the fit using just the important features identified in Fig. 2 are shown. Formally, each of the model predictions is a fully Bayesian posterior prediction. That is, it is the expected similarity of a pair of stimuli averaged across the posterior distributions of the saliency weight and additive constant parameters. For the full feature list analyses, all of the indicator variables are fixed to include all features; for the important feature analyses, just the first features have their indicator variables set for inclusion.

The similarity fits in Fig. 3 support a number of conclusions. First, the fits of the full feature set indicate that our model achieves a basic level of descriptive adequacy. This is important, because it means the inferences our model makes can be taken seriously. There is no reason to trust estimates of model parameters unless the model is able to fit the data on which these estimates are based.

The second observation is that using just the subset of important features, totaling only about one-quarter of all available features, still fits the similarity data well. This is particularly true for the animal datasets.[2] What these analyses show, then, is that using just the important features can provide a very good descriptive account of the similarities between the stimuli, and so can provide a suitable basis for their representation.

4.3. Interpreting feature importance

Tables 3–6 summarize the features found to be the most and least important by our model for each of the four datasets. In each table, the 20 features with the highest importance measures are shown at the top, and the 20 features with the lowest importance measures at the bottom. After each feature label, the importance measure is shown, followed by the pattern of belonging to the stimuli for that feature. Where a stimulus has a feature indicated by a bullet point.

In general, the most and least important features found in Tables 3–6 are intuitively reasonable. The most important features seem to express useful properties of animals or objects, while many of the least important ones do not. For example, Table 3 lists "is a bird", "is an insect", "can fly", "has bones", "lives in water" and so on as important features. Many of these features indicate category membership, while others capture basic properties of the animals. Many of these important features re-appear in Table 4, in the modified context of the animals considered in the second dataset.

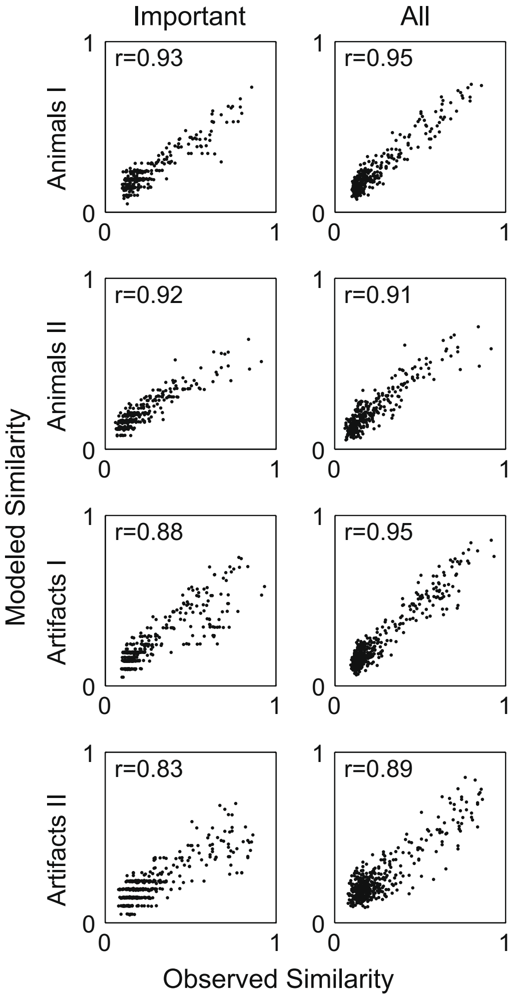

Fig. 3. The fit of the model to the similarity data, for each dataset, and using the most important or all features. Each panel shows the scatterplot relating the modeled and empirical mean similarities between each stimulus pair, and the correlation coefficient.

The least important features in Table 3 suggest that there can be a number of reasons for a feature not being useful to represent stimuli. Some features, such as "neutral scent" and "appears in stories" seem peripheral or obscure. They are the sorts of features that could be produced, as required in the generation task, but do not seem central to understanding or representing animals. Other features, such as "does not sting" relate to more important concepts, but express the absence of an important property, contrary to common features assumptions about similarity (e.g., a monkey and a rabbit are not psychologically much more similar because neither of them sting). Table 3 includes "has a sting" among the important features, expressing the same concept in terms of the presence of the property. We acknowledge, however, that a few of the least important features in Table 3, such as "strong animal" and "is big", seem useful. We believe their lack of importance stems from more idiosyncratic reasons relating to the limited set of animal stimuli in the particular dataset. That is, while these features are probably important in general, they are not for the particular narrow context

[2] The ability of the important feature model to fit the similarity data better than the all features model may seem paradoxical, since the all features model includes the important features model as a special case. It is the fully Bayesian nature of the posterior prediction—based on averaging across posteriors, not maximizing with parameter point estimates—that gives rise to this possibility. Although the all features model could set the relevant weight parameters to zero to mimic the important features model, finite empirical data does not provide the evidence to make this inference, and the uniform priors on the weights continue to affect prediction.

Table 3
The most and least important features for the first animal dataset. (Imp. = Importance).

Feature	Imp.	Mammals					Birds					Fish					Insects					Reptiles				
		Monkey	Rabbit	Sheep	Pig	Bat	Seagull	Stork	Ostrich	Swan	Swallow	Shark	Squid	Carp	Ray	Salmon	Bumblebee	Mosquito	Caterpillar	Spider	Wasp	Boa	Dinosaur	Crocodile	Iguana	Salamander
is a bird	1.00						•	•	•	•	•															
is an insect	1.00																•	•	•	•	•					
can fly	1.00					•	•	•			•						•	•			•					
is a fish	1.00											•	•	•	•	•										
has wings	1.00					•	•	•	•	•	•															
is a reptile	1.00																					•	•	•	•	•
has bones	1.00											•		•	•	•										
has little eyes	1.00					•	•				•		•		•				•			•				
lives in water	1.00									•		•	•	•	•	•								•		•
eats fish	1.00						•	•		•		•	•	•	•	•								•		
has four paws	1.00	•	•	•	•																		•	•	•	•
lives on a farm	1.00		•	•	•																					
has a sting	1.00																•				•					
gives pain	1.00																	•		•	•					
has scales	1.00													•	•	•						•	•	•	•	•
small ears	1.00	•			•	•	•	•	•	•	•															
jumps	1.00	•	•																							
has hoofs	1.00			•	•																					
is irritating	1.00																•	•		•	•					
lives in the wild	0.10	•	•			•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
does not produce sound	0.10		•									•	•	•	•	•			•	•						
lives on land	0.10	•	•	•	•	•	•	•	•	•	•						•	•	•	•	•	•	•	•	•	•
lays eggs	0.09					•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
appears in stories	0.09		•	•	•			•		•		•								•			•	•		
is brown	0.09	•	•						•					•	•					•		•	•		•	•
can be lethal to man	0.08											•								•	•	•	•	•		
is big	0.08	•						•	•			•										•	•	•		
strong animal	0.08	•							•			•										•	•	•		
eats insects	0.08					•	•	•	•	•	•									•		•	•	•	•	•
does not migrate in winter	0.08	•	•	•	•	•	•		•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
lives in Africa	0.07	•	•	•	•			•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
can be dangerous	0.07	•										•								•		•	•	•		
is light	0.07		•								•				•	•	•	•	•	•	•				•	•
lives in the zoo	0.06	•				•		•	•	•		•	•		•	•				•		•		•	•	•
found in warm places	0.06	•							•		•	•								•		•		•	•	•
can kill people	0.06									•		•								•		•	•	•		
does not sting	0.05	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•		•	•	•	•	•
neutral scent	0.05	•	•			•	•	•	•	•	•						•	•	•	•	•	•	•	•	•	•
lives in tropics	0.05	•						•	•			•								•		•		•	•	•

Table 4
The most and least important features for the second animal dataset. (Imp. = Importance).

Feature	Imp.	Mammals					Birds					Fish					Insects					Reptiles				
		Squirrel	Hedgehog	Donkey	Dog	Cow	Eagle	Rooster	Blackbird	Parrot	Peacock	Trout	Eel	Sardine	Whale	Swordfish	Cricket	Dragonfly	Moth	Grasshopper	Butterfly	Alligator	Cobra	Lizard	Frog	Tortoise
is an insect	1.00																•	•	•	•	•					
is a bird	1.00						•	•	•	•	•															
can fly	1.00						•	•	•	•								•	•		•					
has wings	1.00						•	•	•	•	•							•	•		•					
lives in the sea	1.00											•	•	•	•	•										
eats grass	1.00			•		•																				
is not eaten	1.00	•	•				•		•	•	•						•	•	•	•	•		•	•		
lives in a damp climate	1.00											•	•	•	•	•		•				•			•	•
has scales	0.99											•	•	•		•						•	•	•		•
is a fish	0.99											•	•	•		•										
lives on a farm	0.97			•	•	•		•																		
is elongated	0.97											•	•		•	•						•	•	•		
lays eggs	0.96						•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•
can jump far	0.95																•			•					•	
is aggressive	0.94																					•	•			
is a freshwater fish	0.94											•	•									•				
is a reptile	0.93																					•	•	•	•	•
does not live long	0.93																	•	•		•					
has a shield	0.93																•			•						•
is a pet	0.91				•					•																
has small paws	0.11	•	•						•	•												•		•		•
has claws	0.10	•			•		•	•	•	•	•											•		•		
is light	0.10	•	•						•	•		•	•	•		•	•	•	•	•	•			•	•	
is not very big	0.10	•	•					•	•	•		•	•	•			•	•	•	•	•			•	•	•
has a tail	0.10	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•						•	•	•		•
has black eyes	0.10	•	•	•		•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
is edible	0.10	•	•	•	•	•		•				•	•	•	•	•				•		•			•	
has little eyes	0.10	•	•					•	•	•			•	•									•			
does not taste god	0.10	•	•	•	•		•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•
is colored	0.09				•			•		•	•							•			•		•	•	•	
does not sting	0.09	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•
there are many kinds of it	0.08		•	•	•	•		•		•					•						•		•	•	•	•
has a small head	0.08	•	•					•	•	•								•						•	•	•
has a tongue	0.07	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•					•	•	•	•	•	•
smaller than a horse	0.07	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•
occurs frequently	0.06	•	•	•	•	•		•	•			•	•	•		•	•		•	•	•			•	•	
the meat is eaten	0.06		•		•	•	•	•				•	•	•	•	•						•	•		•	
lives in warm countries	0.05	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
is small	0.04	•	•					•	•	•		•	•	•		•	•	•	•	•	•		•	•	•	
lives in Africa	0.04	•		•	•	•	•	•		•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

Table 5
The most and least important features for the first artifact dataset (Imp. = Importance).

Feature	Imp.	Clothes					Kitchen					Musical					Tools					Vehicles					Weapons				
		Pants	Jeans	Scarf	Sweater	Bathing suit	Bottle	Oven	Grater	Apron	Fork	Bass guitar	Flute	Harmonica	Piano	Tambourine	Hammer	Knife	Plane	Nail	File	Bicycle	Hovercraft	Hot air balloon	Tram	Train	Axe	Rifle	Spear	Stick	Sword
feels rough	1.00								•										•		•										
is a vehicle	1.00																					•	•	•	•	•					
is worn on legs	1.00	•	•																												
is made of fabric	1.00	•	•	•	•	•				•																					
is a tool	1.00								•								•	•	•	•	•						•				
is a musical instrument	1.00											•	•	•	•	•															
is small	1.00										•	•					•		•	•								•	•		
is ironed	1.00	•	•		•					•																					
damages delicate surfaces	1.00								•		•						•	•	•	•	•						•		•		•
has a grip	1.00							•	•		•						•	•	•		•							•			•
is public transport	1.00																								•	•					
used to threaten	1.00						•				•						•	•									•	•	•	•	•
has wheels	1.00																					•			•	•					
can injure	1.00										•						•	•	•	•					•	•	•	•	•	•	•
used with your hands	1.00						•	•	•		•	•	•	•	•	•	•	•	•	•	•						•	•	•	•	•
made of fibers	1.00	•	•	•	•	•				•														•							
used in the kitchen	1.00						•	•	•	•	•							•													
can cause injury	1.00																•	•	•	•	•						•	•	•	•	•
is silver-colored	1.00								•		•						•	•	•	•	•						•				•
can break things	0.99																•	•			•						•	•	•	•	•
is easy to work with	0.05						•	•	•	•	•						•	•	•	•	•						•				
used on a yard	0.05																•	•	•	•							•				
costs a lot of money	0.05							•				•	•		•								•	•	•	•					•
is hard	0.04	•					•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•
is for all ages	0.04	•		•	•	•	•	•	•	•	•	•	•	•	•	•						•			•	•					
made of different tissues	0.04	•		•	•	•	•		•	•	•																				
is strong	0.04							•									•	•		•			•		•	•	•				•
both large and small	0.04						•	•	•		•		•		•	•	•	•	•	•	•	•	•				•	•		•	
seen on television	0.04	•	•	•	•	•						•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
is useful	0.04			•		•	•	•	•	•	•						•	•	•	•	•	•	•		•	•					
gone through tech revolution	0.04							•				•			•								•		•	•		•			
comes in very handy	0.04					•	•	•	•	•	•						•	•	•	•	•	•			•	•	•			•	•
shines	0.04										•		•	•				•		•		•						•			•
is big or small	0.03						•	•							•		•	•		•		•					•			•	
is expensive	0.03							•				•			•								•	•	•	•		•			•
made of metal	0.02						•	•	•		•	•	•	•	•			•	•	•	•	•	•		•	•	•	•			•
is light	0.02	•		•	•	•	•			•	•		•			•		•	•	•											
often used	0.02						•	•	•		•						•	•		•		•		•	•						
is durable	0.02							•			•						•	•	•	•	•		•		•	•	•	•			•
is stainless	0.02						•	•	•		•							•		•		•	•		•	•	•		•		•

Table 6
The most and least important features for the second artifact dataset (Imp. = Importance).

Feature	Imp.	Clothes					Kitchen					Musical					Tools					Vehicles					Weapons				
		Shirt	Hat	Costume	Boots	Skirt	Furnace	Fridge	Mixer	Scissor	Sieve	Accordion	Drum set	Harpsichord	Trumpet	Violin	Chisel	Crowbar	Wheelbarrow	Grinding disc	Vacuum cleaner	Motorbike	Sled	Taxi	Tractor	Airplane	Bow	Dagger	Grenade	Pistol	Shield
is a garment	1.00	•	•	•	•	•																									
is a vehicle	1.00																		•			•	•	•	•	•					
produces music	1.00											•	•	•	•	•															
is a machine	1.00						•	•	•											•	•										
damages delicate surfaces	1.00									•							•	•		•								•			
used to make war	1.00																									•	•	•	•	•	•
has sleeves	1.00	•		•																											
costs money	1.00	•	•	•	•	•						•	•	•	•	•								•		•					
seen on television	0.99	•	•	•	•	•						•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
made of fabric	0.99	•	•	•		•																									
sometimes dirty	0.99				•														•						•						
is white	0.99						•	•																							
is worn if warm	0.99	•				•											•	•	•	•											
used by workers	0.99																•	•	•	•											
can be used as weapon	0.98									•								•		•							•	•	•	•	
can be played	0.98											•		•	•	•															
is made of cotton	0.98	•		•		•																									
used in folk music	0.98											•				•															
has a steering wheel	0.98																					•		•	•	•					
is used to cut	0.98									•																		•			
can break	0.12													•		•							•					•		•	
produces different sounds	0.12											•	•	•	•	•								•	•	•					
is functional	0.12						•	•	•	•	•						•	•	•	•	•	•		•	•	•	•	•	•	•	•
exists in different materials	0.12	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•		•		•	•	•	•	•	•	•	•		•	
is hard	0.11						•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
is durable	0.11						•	•									•	•	•	•	•				•	•		•		•	•
has a grip	0.10							•	•	•	•						•		•			•					•	•		•	•
is a machine	0.10						•	•	•											•	•	•		•	•	•					
exists in different colors	0.10	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•
produces a lot of sound	0.10											•	•		•					•	•	•			•	•			•	•	
is strong	0.10						•	•									•	•	•	•		•		•	•	•		•	•	•	•
is dangerous for children	0.08						•		•	•								•		•		•		•		•	•	•	•	•	
exists in different lengths	0.08	•			•	•								•		•	•	•					•	•	•	•	•	•	•	•	•
constitutes a whole	0.08	•	•	•		•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
needs to be cleaned	0.08				•		•	•		•					•		•					•	•	•	•	•	•	•		•	•
can sink	0.07	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•	•	•	
is for all ages	0.07				•	•	•	•	•	•	•	•	•	•	•	•							•	•		•					
is stainless	0.06						•	•	•	•	•				•		•			•		•		•	•	•		•		•	
produces noise	0.06							•	•			•	•	•	•	•				•	•	•		•	•				•	•	
makes a sound	0.06						•	•	•			•	•	•	•	•				•	•	•			•				•	•	

in which empirical data are available in the specific dataset. Some evidence for this belief is provided by the lack of re-appearance of features like "is big" in Table 4.

Broadly similar insights are suggested by the lists of most and least important features for the artificial domain datasets in Tables 5 and 6. The most important features make intuitive sense, and often provide information about categories. Many of the least important features seem unusual or peripheral, and there is more consistency across the individual datasets for the concepts expressed by the more important features. All of these sorts of conclusions, however, are heavily subjective, and open to debate. In addition, the decision to show only the 20 most and least important features in the tables has no principled basis, and it is not clear if the conclusions would be affected by considering all of the features and their measures of importance. For these reasons, the results in Tables 3–6 are intended to suggest analyses, rather than support firm conclusions. In fact, there are a number of obvious and potentially important analyses suggested by the details in the tables, which we pursue next. Visually, for example, it seems that more important features seem to belong to fewer objects, and more closely adhere to category boundaries. In the next section, we report to results of a series of quantitative analyses, based on the information represented by Tables 3–6, that attempts to test these sorts of hypotheses more precisely.

5. Understanding feature importance

The information the model provides about which features are likely to be important for representing stimuli allows us to consider basic theoretical issues about what makes for a good feature. The additional information in the Leuven Concept Dataset relating each stimulus to a category means we can address this question in the context of how more or less important features relate to category structures. In this section, we present some analyses tackling these issues.

The relationship between stimulus similarity and category structures and category learning has been extensively studied in cognitive psychology (Ashby & Lee, 1991; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984, 1986, 1989, 1991; Rips, 1989; Sloutsky & Fisher, 2004; Smith & Sloman, 1994). Among the most relevant theorizing and empirical findings for our study are measures of feature structure developed to explain basic level categorization. Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) introduced the idea that people prefer to categorize objects at a particular level in a hierarchy of categories, which they called the basic level, and postulated this preference arises from natural patterns of feature co-occurence in the world. Subsequently, a number of measures of feature structure have been proposed to explain basic level categorization (e.g., Corter & Gluck, 1992; Gosselin & Schyns, 2001; Jones, 1983; Kloos & Sloutsky, 2006; Mervis & Rosch, 1981). At their core, many of these measures share two complementary concepts: cue validity, the conditional probability of an object belonging to a category c given it has a feature f, $p(c|f)$, and category validity, the conditional probability of an object possessing feature f given that it belongs to category c, $p(f|c)$. Neither measure is singly sufficient to predict basic level categorization (see Medin (1983) and Murphy (1982), for arguments against cue validity and category validity, respectively), which has led to vigorous debate over the appropriate way to combine the two (e.g., Corter & Gluck, 1992; Jones, 1983; Medin, 1983; Mervis & Rosch, 1981).

Our analyses proceed from the working hypothesis that a features importance is related to its role in forming categories among the stimuli. The literature on basic level categorization then suggests examining quantities related to cue and category validity. We look at four such quantities: feature density, category likelihood ratio, category density, and category span. Feature density and category span measure different aspects of cue validity, since

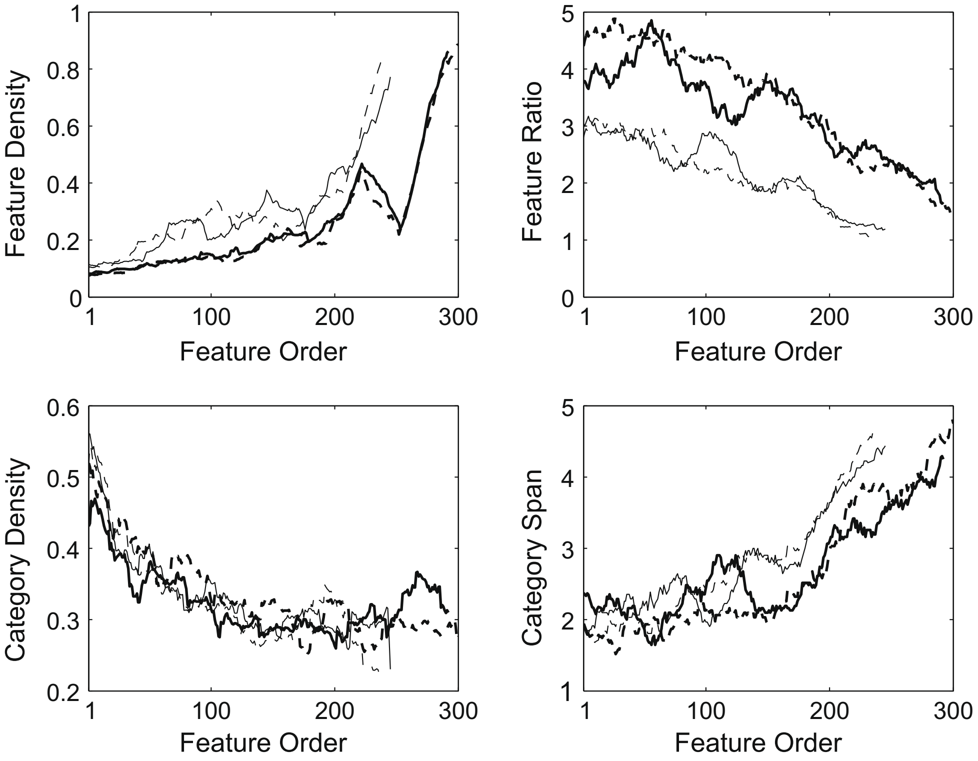

Fig. 4. The pattern of change in four measures—feature density, feature ratio, category density and category span—over features ordered from most to least important. In each panel, the thicker lines correspond to the animal datasets, while the thinner lines correspond to the artifact objects datasets. Solid lines correspond to the first dataset in each case, while broken lines correspond to the second dataset.

features with high validity should both be fairly rare and particular to a small number of categories. The category likelihood ratio is closely related to category validity. Finally, we compute a measure combining cue and category validity, category density (Kloos & Sloutsky, 2006). Our first four results are summarized in Fig. 4.

5.1. Feature density

The first measure we consider is feature density, which is simply the proportion of all possible stimuli that have a feature. This measure can vary from low values near zero, for features that belong to very few stimuli, to large values near one, for features belonging to almost every stimulus. The top-left panel in Fig. 4 shows the pattern of change in feature density, as the features progress from most to least important. Each line in the figure corresponds to one of the four datasets, with the animals datasets shown by the thicker lines. The curves have been smoothed for legibility, by averaging over a small window.

The clear result in Fig. 4 is that important features likely to be used in representing stimuli tend to have low feature density, and so belong to relatively few stimuli. The most important features belong to about 10% of the possible stimuli, and so are rare or sparse, whereas the least important features belong to 80% or more of the stimuli, and so are very broad or common. This is not a surprising result, especially given the assumption of a common-features model of similarity. Under this model, sparse features are ideal for identifying narrow or niche aspects of small sets of stimuli that provide an important contribution to their mutual similarity. Broad features, in contrast, will increase the similarity of most stimuli, which is a role already covered (approximately) by the additive constant in the similarity model. The basic message of the analysis of feature density is that important features tend to belong to few stimuli.

5.2. Category likelihood ratio

The second measure we consider involves both feature and category information, and is called the category likelihood ratio. The basic idea is to measure whether a feature is more likely to belong to stimuli if those stimuli are in the same category. Formally, the likelihood ratio measure compares the probability that two stimuli chosen at random from within a category will both have a feature, to the probability that two stimuli chosen at random will both have the feature. The results in the top-right panel of Fig. 4 show that, for the most important features in the ordering, it is about 3–5 times more likely for the feature to be common within rather than across categories. This likelihood decreases towards the chance base-line of 1 as the features become less important.

It is interesting to note that the animal and artifact domain datasets are different in their behavior on this measure, especially in the sense that the limiting ratio of 1 is basically achieved for the artifact domain, but not for the animal domain. This suggests that people are able to generate more features that are sensitive to category boundaries for animals than artifacts, or, equivalently, that the category boundaries are sharper or more pronounced for the animals. These details aside, however, the important characteristic of the likelihood ratio measure is that it decreases as features become less important. The basic message of the category likelihood ratio analysis is that important features are more likely to co-occur for stimuli that belong to the same category.

5.3. Category density

The third measure we consider is a natural variant of the category density measure developed by Kloos and Sloutsky (2006), who used it to explore developmental issues in the ease of learning categories, and the use of similarity- versus rule-based category learning systems. Category density is similar to the category likelihood ratio, in the sense that it compares how features co-occur within categories against their overall patterns of co-occurrence across all categories. The difference is that, rather than relying on likelihood ratios, the category density measure uses the notion of entropy from information theory (Cover & Thomas, 2008) to assess within versus between category feature structure. Formally, if the kth feature belongs to stimuli in the mth category with probability p_{km}, and to all stimuli with overall probability p_k, the category density measure we use is given by $1 - (\sum_m p_{km} \log p_{km})/(p_k \log p_k)$. This is a variation on the measure used by Kloos and Sloutsky (2006), who included both the presence and absence of features in their calculations. We focus on only the presence of features in determining entropy, consistent with our assumption of a common-features similarity model. In both variants of the measure, however, larger values of category density correspond to features that provide more information about how stimuli belong to categories.

The bottom-left panel of Fig. 4 shows the pattern of change of category density over the ordered features. It is clear that, for all datasets, the category density is higher for more important features at the beginning of the order. This indicates that important features tend to contain information about the category structure of a set of stimuli, identifying pairs of stimuli that belong to the same category. Qualitatively, this is the same finding as we obtained from the category likelihood ratio measure. It is interesting to note, however, that the category density measure does not show the separation between the animal and artifact datasets evident in the likelihood ratio measure. Nevertheless, the basic message of the category density measure is that features carrying information about category structure are more likely to be used to represent stimuli.

5.4. Category span

The final measure in Fig. 4 is the category span of the features. This is simply the number of categories containing at least one stimulus that had the feature. As can be seen in the bottom-right panel of Fig. 4 this measure starts around 2 for the most important features, but increases towards the maximum of 5 (for the animal datasets) or 6 (for the artifact datasets) as features become less important. In this way, the category span measure provides another form of evidence for the conclusion that people are most likely to use features that tell them about categories.

5.5. The size principle

Tenenbaum and Griffiths (2001) described an interesting and insightful relationship between features and their salience that they called the size principle, and is also considered by Navarro and Perfors (2010) and Steyvers (2010). This principle follows from their basic theorizing about how the core cognitive capability of generalization operates over structured hypothesis spaces for representing stimuli. The basic claim is that more specific features belonging to fewer stimuli (i.e., low feature density in our terminology) should be given greater salience. Tenenbaum and Griffiths (2001) present a basic empirical confirmation of this idea by showing that the feature weights found in additive clustering representations of similarity data decrease as the density of the features increases.

Our modeling automatically permits the same empirical check, which is presented in Fig. 5. Each panel shows the relationship between feature density and feature weight. The rows correspond to the four datasets, while the columns correspond to the important feature and all feature analyses. It is clear from Fig. 5 that a pattern

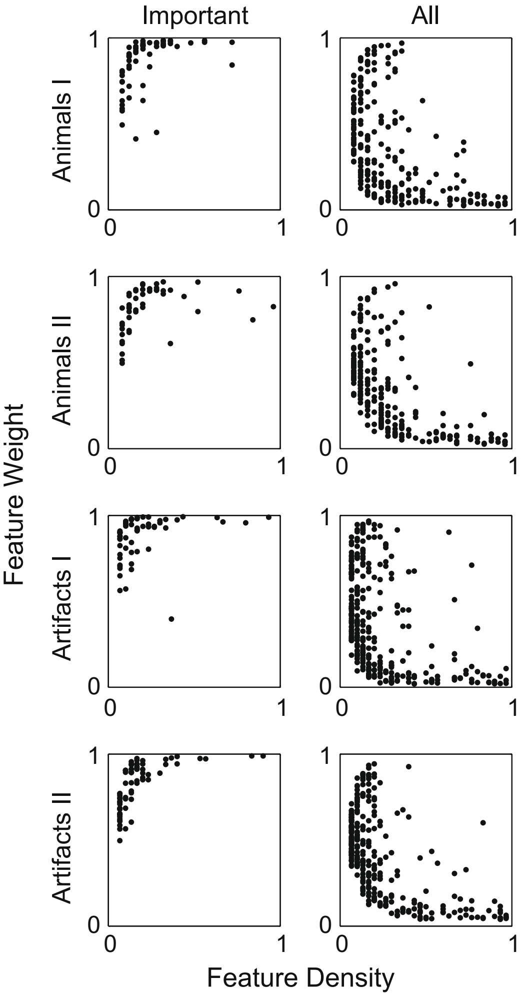

Fig. 5. The relationship between feature density and feature weight for each dataset, using all of the features, and just the subsets of important features.

highly consistent with the size principle is observed when all of the features are considered. Features with low density (i.e., belonging to few stimuli) tend to have high expected feature weights, but features with high density tend to have much lower weights. It is interesting to note, however, that the same basic relationship does not hold when just the subsets of important features are considered. Rather, feature density is now positively correlated with feature weight, with the few high density features included among the important features invariably having very high feature weights.

We do not necessarily interpret this as strong evidence against the size principle. It seems likely people have access to most of the features in the full list. Many of these will involve many stimuli, and will not deserve large weight in the overall representation. Often, the features that will need high weights will involve just a few stimuli, and allow fine-grained distinctions to be made. This combination of low density but high feature weight is consistent with the size principle.

But our current analyses also suggest that restricting the features to just those needed to account for similarity comparisons selects only a few of the most useful high density features. These serve the purpose of establishing basic gross distinctions across all of the stimuli (e.g., whether an animal is large or small), and so also have high weight in assessing similarities. This combination of high density and high feature weight is not consistent with the size principle, and is a result of the focus of the model on finding a small set of features for representation.

6. Conclusion

One of the most basic challenges in cognitive science is to understand how people represent stimuli. One of the most useful and most popular answers, especially when thinking about higher-order cognitive processes, is that stimuli are represented in terms of the presence or absence of a set of meaningful features. In this paper, we have developed a model for finding feature representations based on feature generation and similarity judgment tasks, and applied them to the Leuven Concept Database.

Methodologically, the model has a number of strengths, including the ability to identify the importance of generated features for the core cognitive capability of assessing similarities, and preserving the semantic labels associated with each feature to assist the interpretation of a representation.

Theoretically, our modeling permitted a range of explorative analyses aimed at understanding what makes a feature a useful or important part of representation. Our results suggested that features that belong to relatively few stimuli, identifying what makes them similar, and especially identifying the categorical structures within which stimuli are organized, tend to be the more important ones.

There are a number of obvious and straightforward ways in which our modeling approach can be generalized. Most fundamentally, our current model relies on specific assumptions about how the processes of feature generation and similarity judgment work. Extending or changing these assumptions will lead to new and potentially better models. For feature generation, it seems plausible that some key features might be omitted, and so it could be useful to allow for the inference of latent features in addition to those observed. For similarity judgments, there are a range of more involved accounts than additive common features worth exploring (e.g., Navarro & Lee, 2004). One particularly promising avenue in this regard involves proposing a set of relational types describing how features can interact to affect similarity, and so generalizing existing common and distinctive features models.

All of these future possibilities could be implemented within our basic graphical modeling framework, and the Leuven Concept Database would continue to provide an ideal source of empirical data. We are sure we have just scratched the surface of what the Leuven data can tell us about the features people use, and why they use them.

References

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120*, 150–172.
Bush, R. R., & Mosteller, F. A. (1951). A model for stimulus generalization and discrimination. *Psychological Review, 58*, 413–423.
Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin, 111*(2), 291–303.
Cover, T. M., & Thomas, J. A. (2008). *Elements of information theory*. New York, NY: Wiley.
De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M., Voorspoels, W., et al. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods, 40*(4), 1030–1048.
Dennis, S. J., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452–478.
Eisler, H., & Ekman, G. (1959). A mechanism of subjective similarity. *Acta Psychologica, 16*, 1–10.
Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1*(3), 515–534.
Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650–669.
Gosselin, F., & Schyns, P. G. (2001). Why do we SLIP to the basic level? Computational constraints and their implementation. *Psychological Review, 108*(4), 735–758.
Gregson, R. A. M. (1975). *Psychometrics of similarity*. New York: Academic Press.
Hintzman, D. L. (1984). Minerva-2 – A simulation-model of human-memory. *Behavior Research Methods Instruments and Computers, 16*(2), 96–101.
Jones, G. V. (1983). Identifying basic categories. *Psychological Bulletin, 94*, 423–428.
Jordan, M. I. (2004). Graphical models. *Statistical Science, 19*, 140–155.

Kloos, H., & Sloutsky, V. M. (2006). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General, 137*(1), 52–72.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22–44.

Lee, M. D. (2002). Generating additive clustering models with limited stochastic complexity. *Journal of Classification, 19*(1), 69–85.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin and Review, 15*(1), 1–15.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin and Review, 9*(1), 43–58.

Love, B. C., Medin, D. L., & Gureckis, T. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

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.

Medin, D. L. (1983). In B. Shepp & T. Tighe (Eds.), *Interaction: Perception, development, and cognition* (pp. 203–230). Hillsdale, NJ: Erlbaum.

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

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89–115.

Murphy, G. L. (1982). Cure validity and levels of categorization. *Psychological Bulletin, 91*, 174–177.

Navarro, D. J., & Griffiths, T. L. (2004). Latent features in similarity judgment: A nonparametric Bayesian approach. *Neural Computation, 20*(11), 2597–2628.

Navarro, D. J., & Lee, M. D. (2004). Common and distinctive features in stimulus similarity: A modified version of the contrast model. *Psychonomic Bulletin and Review, 11*(6), 961–974.

Navarro, D. J., & Perfors, A. F. (2010). Similarity, feature discovery and the size principle. *Acta Psychologica, 133*(3), 256–268.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*(1), 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception and Psychophysics, 45*, 279–290.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23*, 94–140.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). *The adaptive decision maker*. New York: Cambridge University Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Restle, F. (1959). A metric and an ordering on sets. *Psychometrika, 24*(3), 207–220.

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

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.

Shepard, R. N., & Arabie, P. (1979). Additive clustering representations of similarities as combinations of discrete overlapping properties. *Psychological Review, 86*(2), 87–123.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*(2), 145–166.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166–188.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition, 22*, 377–386.

Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychologica, 133*(3), 234–243.

Tenenbaum, J. B. (1996). Learning the structure of similarity. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.). *Advances in neural information processing systems* (Vol. 8, pp. 3–9). Cambridge, MA: MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629–640.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

Zeigenfuse, M. D., & Lee, M. D. (2008). Finding feature representations of stimuli: Combining feature generation and similarity judgment tasks. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 1825–1830). Austin, TX: Cognitive Science Society.