PUBLISHED VERSION

Stephens, Rachel Gloria; Navarro, Daniel Joseph One of these greebles is not like the others: Semi-supervised models for similarity structures Proceedings of the 30th Annual meeting of the Cognitive Science Society, 23-26 July, 2008: pp.1996-2001

© the authors

PERMISSIONS

correspondence from:
Business Mgr
Cognitive Science Society Inc. [cogsci@psy.utexas.edu]
University of Texas - Austin
Department of Psychology
108 E. Dean Keeton, Stop A8000
Austin

The copyright for articles and figures published in the Proceedings are held by the authors, not the Society

http://hdl.handle.net/2440/54275

One of These Greebles is Not Like the Others: Semi-Supervised Models for Similarity Structures

Rachel G. Stephens (rachel.stephens@adelaide.edu.au)
School of Psychology, University of Adelaide
SA 5005, Australia

Daniel J. Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide
SA 5005, Australia

Abstract

When studying human concepts, artificial categories are often used that are very simple, with a chosen set of discrete features. Images such as "greebles", animal-like three-dimensional figures, are a more realistic and interesting alternative artificial stimulus set for categorization experiments. However, especially for more complex stimuli, it is not obvious that participants' perceived similarity structure will be as simple as the structure used to construct the stimuli. Therefore, it is safest to first empirically obtain a measure of the stimuli's structure. We demonstrate the use of an odd one out task in order to model people's mental representations of a set of greebles, an example set of moderately complex stimuli. We show that, although the greebles were constructed such that they can be classified into two sets of categories, these categories have a considerable amount of additional latent structure that can be extracted by similarity modeling.

Keywords: greebles; similarity; categorization.

Consider a simple "odd one out" problem, in which people are asked to decide which out of a collection of items is least like the others. It is a fairly natural task, and in real life it is sometimes used to help people acquire novel concepts - it is, for instance, one of the recurring segments in the U.S. children's TV show *Sesame Street*. In the children's show, only one salient characteristic is typically varied across the presented items, but this need not always be the case. In a more general setting, picking an odd one out will tend to pit different aspects of the stimuli against one another, potentially revealing latent mental representations without resorting to somewhat unnatural tasks like rating similarity. Experimentally, the idea is based on the kind of relative judgment tasks that are common in psychophysics (e.g., Woodsworth & Schlosberg, 1954, ch. 20), but less common when measuring similarities (but see, e.g., Navarro & Lee, 2002).

In this paper, we pursue this idea, investigating how people make choices involving "greebles", complex visual stimuli that have been used as a set of relatively homogenous non-face objects in a range of experimental tasks (e.g., Gauthier & Tarr, 1997; Gauthier, Williams, Tarr, & Tanaka, 1998; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002). We first explain why it is useful to empirically uncover people's representations of complex stimuli such as greebles. We provide a brief overview of the greeble stimuli, and the manner in which they have been used in previous research. We then discuss the kinds of structured representations that might account for people's decisions about greebles in an odd one out task. Having done so, we present experimental data that illustrate that the simple classification system generally used to describe greebles is insufficient, and use the models developed to infer a richer representation of greebles.

Why the Greebles?

Research on category learning and category-based inferences usually involves selecting category members on the basis of some similarity structure. For example, a "family resemblance" structure is often used, where no one feature defines a category, but all members have some features in common, such as color, size or shape. Often researchers use artificial categories so that they can control the relevant category knowledge available to participants. In order to control the similarity within and between categories, stimuli are often constructed using discrete features, which may be presented to participants as written lists of features (e.g., Rehder, 2006) or simple stimuli such as two-dimensional fictitious insects or geometric shapes (e.g., Johansen & Kruschke, 2005; Yamauchi, Love, & Markman, 2002).

One difficulty with this process is that people may not perceive stimuli in terms of the features that the experimenter thinks are most obvious; and if they do, it is almost certain that they do not weight features as equally important. Even in toy domains involving simple colored shapes the latter assumption is typically violated, though the violations are typically minor (e.g., Lee & Navarro, 2002, table 5). A bigger concern is the lack of ecological validity – if what we ultimately want to know about is category-based reasoning for complex real-world categories, such simple and abstract experimental stimuli may be too far removed from real categories. Furthermore, during an experiment, participants may be less motivated to reason carefully about simple figures or feature lists than stimuli that are more engaging and realistic. With this in mind, the alternative is to work with naturalistic stimulus domains. The awkward problem in this context is that the underlying mental representations are even harder to specify in such contexts, and a priori knowledge can vary considerably across people.

In view of these difficulties, moderately complex artificial stimuli such as the animal-like three-dimensional "greebles" (described below) offer an attractive alternative for use in categorization studies. The greebles are more interesting and realistic than many artificial stimuli that are used. However, with more complex stimuli such as greebles, it is difficult to tell a priori which "features" participants may consider if asked to reason about the stimuli. As with most artificial stimuli, we have some relevant prior knowledge about how the greebles are likely to be represented, but cannot be sure a priori which features are more salient, or whether there are subtle relationships between them. Accordingly, if such items are to be used, it is safest to first empirically obtain a measure of the stimuli's structure. In this paper, we demonstrate the use of an odd-one-out task for the greebles, and apply some simple "semi-supervised" models to the resulting similarity data.

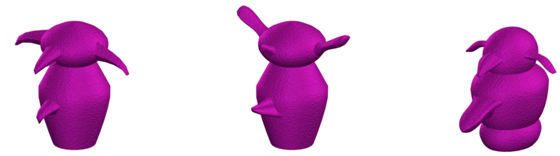

Figure 1: Three greebles categorized according to two different systems, gender and family. The first two greebles are in family 1, while the rightmost one belongs to family 2. The left and right greebles are female, and the middle greeble is male.

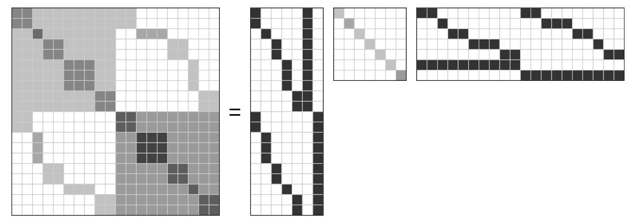

Figure 2: An example of a similarity structure induced by an uncorrelated category system, shown as a graphical representation of the equation $\mathbf{S} = \mathbf{Z}\mathbf{W}\mathbf{Z}^{\mathsf{T}}$. The binary matrices $\mathbf{Z}$ are defined by the categorization of the greebles into families and genders, but the diagonal weight matrix $\mathbf{W}$ is unknown.

Meet the Greebles

The greeble images that we consider in this experiment consist of a set of artificial three-dimensional "animals", made available courtesy of Michael Tarr (http://www.tarrlab.org/), and illustrated in Figures 1 and 5. All greebles have a main body and four protruding appendages arranged in a roughly similar configuration, though the shapes of appendages vary between individuals. The greebles are constructed such that they belong to one of five *families* determined by their body shape, and one of two *genders*, indicated by the direction in which the appendages point (upward = males) (Gauthier & Tarr, 1997). All greebles have a textured purple shading, with two viewpoints or camera angles available.[1]

Although the greebles are interesting visual stimuli in their own right, they have been used primarily to contest the hypothesis that there are mechanisms specialized for recognizing upright human faces. The idea is that greebles are nonface objects constructed systematically from similar features, and are hence endowed with a degree of visual homogeneity similar to that seen with faces. The basic approach (e.g., Gauthier & Tarr, 1997) is to train people to become "greeble experts" via an identification task (i.e., learning to name each greeble), and then test whether putatively face-specific effects also occur with greebles once this expertise is acquired. Demonstrated effects include particular patterns or locations of brain activity (Gauthier et al., 1999; Rossion et al., 2002), or behavioral effects such as sensitivity to configural information (Gauthier & Tarr, 1997; Gauthier et al., 1998).

It is not unreasonable to suppose that, like faces, greebles are complex and interesting perceptual objects. Nevertheless, previous research has tended to rely on (at least implicitly) a simple description of greebles in terms of the gender and family membership of each image (e.g., Behrmann, Marotta, Gauthier, Tarr, & McKeeff, 2005; Gauthier et al., 1998). For instance, these two category systems have been relied upon to control for the similarity among groups of greebles, or in constructing composite greebles. Moreover, the five families are sometimes assumed to be equally similar to one another, though this seems unlikely. In short, while the procedure used to generate the greebles provides some information about how people perceive them, it is not a complete account.

With this in mind, when substituting greebles for faces, it would be useful to know something about the inherent perceptual relationships among greebles, since this can affect the decisions or performance of participants in a variety of ways. For example, participants make fewer errors while learning to classify simple novel stimuli if the members within each category are very "similar" (Posner & Keele, 1968). When explicit category labels are involved, the general pattern is for more errors to occur "near" a category boundary (Nosofsky, 1988). Since similarity and distance are notoriously difficult to specify a priori, a safe course would be to obtain these measurements empirically. However, no such measures appear to exist for the greebles at present, so one goal of this paper is to provide them.

[1] Throughout this paper we refer to each gender and each family as constituting a category. Obviously, however, one could argue that "upward-pointing" is a feature possessed by several greebles. Mathematically, there are no differences between the two perspectives.

Semi-Supervised Similarity Models

The task of inferring the latent semantic structure to a set of objects is a central one to understanding human intuition and judgment. Whether learnt from a set of similarity relations (Torgerson, 1958; Shepard & Arabie, 1979), a natural language corpus (Landauer & Dumais, 1997; Griffiths & Steyvers, 2002) or any other of a range of possibilities, the key idea is to assume that people rely on rich mental representations of the objects to guide decisions about those objects. In this case, which is not an uncommon one when dealing with artificial-but-plausibly-complex stimuli, the greebles are explicitly constructed to be partitioned in multiple ways (i.e., *gender* and *family*). We refer to each of these ways of classifying the greebles as a "classification system", and note that these systems are known in advance. What is less clear, however, is whether a simple partition is sufficient to describe the system: for instance, if the members of greeble family 1 appear perceptually more similar to family 5 than to family 4, we would need to augment these simple classification systems in some manner. It is to this topic that we now turn.

In what follows outline a minimum amount of formal theory required to model the kind of data that emerge from the odd one out task. We provide a moderately detailed discussion of models that may be used to uncover latent perceptual structure implied by the choice data. Since some aspects

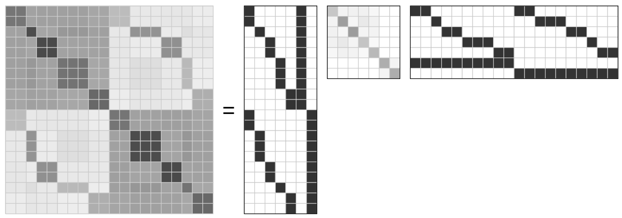

Figure 3: Similarity structure produced by a correlated category system. The binary matrices $\mathbf{Z}$ are defined by the categorization of the greebles into families and genders, but the block diagonal weight matrix $\mathbf{W}$ is unknown. Notice that, while the gross structure is very much the same as that seen in Figure 2, the overall matrix has a "patterned" look.

of the relevant structure (the categories) are known a priori, but other aspects are unknown (relationships between categories), we refer to the models as *semi-supervised similarity models*. Throughout the discussion, we use the following notation for similarities and categories: $\mathbf{S} = [s_{ij}]$ denotes an $n \times n$ matrix of pairwise similarities (where n denotes the number of objects), and $\mathbf{Z} = [z_{ik}]$ is a $n \times m$ matrix of category assignments, where $z_{ik} = 1$ if the ith stimulus belongs to the kth category. However, since $\mathbf{Z}$ is actually composed of two different partitions (one corresponding to gender and one to family, it is convenient to keep these submatrices distinct. In general, if there are Q different systems, then we would let $\mathbf{Z}^{(q)} = [z_{ik}^{(q)}]$ denote the submatrix corresponding to the qth classification system and write $\mathbf{Z} = [\mathbf{Z}^{(1)} \ldots \mathbf{Z}^{(Q)}]$. Accordingly, $z_{ik}^{(q)} = 1$ if stimulus i belongs to the kth category *within system* q.

Uncorrelated categories

The simplest model we consider is one in which we assume that $\mathbf{Z}$ provides all the relevant structure, and use $\mathbf{S}$ only to determine how much importance people apply to each category (the matrix $\mathbf{Z}$ for the greebles is shown in Figure 2). That is, we assume that greebles that belong to the same category will be somewhat similar to each other, but do not assume any other relationships. Formally, we assume that each category represents a fixed (but unknown) degree of association $w_k^{(q)}$ among its members: the predicted similarity between stimulus i and stimulus j is given by

$$s_{ij} = c + \sum_q \sum_k w_k^{(q)} z_{ik}^{(q)} z_{jk}^{(q)}. \tag{1}$$

We would then optimize the weights $w_k^{(q)}$ so that the predicted similarities matched the observed ones as closely as possible. Note that in this case, the empirical similarities are used only to estimate the "salience" of each category of greebles. No additional relationships between the greeble categories are learned.

In essence, though the interpretation of the various terms is slightly non-standard since (a) we are calling each family a "category" rather than a "feature" and (b) we sort the categories into multiple systems, this simple model is mathematically equivalent to the common features model (Tversky, 1977; Shepard & Arabie, 1979). In matrix notation, the model can be written $\mathbf{S} = \mathbf{Z}\mathbf{W}\mathbf{Z}^{\top}$. In this expression, $\mathbf{W}$ is a diagonal matrix whose non-zero elements contain the weights associated with each category. This is illustrated in Figure 2.

Correlated categories

Although the cross-classified model described in Eq. 1 is almost identical to the standard common features model for similarity, the fact that each category belongs to a particular classification system suggests that there are likely to be relationships between some rows of $\mathbf{Z}$. For instance, some *families* of greebles might be more similar to each other. Accordingly, there should be some similarity between items belonging to different categories within the same system. Formally, this induces correlations between the categories, as follows:

$$s_{ij} = c + \sum_q \sum_k \sum_l w_{kl}^{(q)} z_{ik}^{(q)} z_{jl}^{(q)} \tag{2}$$

where $w_{kl}^{(q)} = w_{lk}^{(q)}$ denotes the association between the kth and lth categories in system q. This model can also be written using the $\mathbf{S} = \mathbf{Z}\mathbf{W}\mathbf{Z}^{\top}$ matrix form, but the weight matrix now has a *block diagonal* structure to it (rather than the diagonal structure as per the uncorrelated model), as illustrated in Figure 3. This correlated-category model has no analog in the similarity literature, but is not unlike allowing correlated factors in a factor analysis models. The rationale in both cases is similar – the underlying representational units (features, factors) may not be entirely independent of one another, so the models need the capability to express this.

Structured categories

Although the uncorrelated-category representations are probably too constrained, there is a sense in which the correlated model introduced in the previous section is *too* flexible. To see this, note that the correlated model is formally equivalent to a situation in which we introduce a new "pseudo-category" for every *pair* of old categories that belong to the same system, and create a new free parameter for each one. Accordingly, we also consider a third possibility, in which the category correlations are expressed more parsimoniously, in terms of a set of structured relationships between the various categories. Specifically, we consider the possibility that the categories are organized into a hierarchy, as well as the possibility that the categories are related by a set of overlapping latent features.

Hierarchical structure. Inspired by the additive tree framework used to derive a set of hierarchically-organized relationships (Sattath & Tversky, 1977) from empirical correlations, one approach is to assume a structured representation among categories based on trees. Supposing that we have some tree topology $\mathcal{T}^{(q)}$ that describes the relationships among the categories, and $\ell_{lk} = [\ell_{plk}]$ denote the lengths of edges separating category k from category l, then the various off diagonal weights $w_{kl}^{(q)}$ (for $k \neq l$) are constrained by

$$w_{kl}^{(q)} \propto 1 - \sum_p \ell_{plk}. \tag{3}$$

It is then straightforward to use this expression[2] to construct an optimal tree using standard methods (see Navarro, 2003, for an overview of the range of methods and selection criteria), in the sense of choosing the tree that minimizes the value of the Bayesian information criterion (BIC), in order to find a satisfactory compromise between data-fit and model complexity.

Latent feature structure. It should be noted, however, that tree topologies are quite restricted, and despite their theoretical appeal they may not provide the best model for the kinds of structure observed in many contexts. A natural alternative is to assume that the category correlations can be explained via a set of latent features, such that

$$w_{kl}^{(q)} \propto \sum_b v_b f_{kb} f_{lb}. \tag{4}$$

This approach, which is more closely related to classic featural models (Shepard & Arabie, 1979) assumes that latent structure is implicated in a particular set of categories, but that this structure is not hierarchically organized. Instead, the latent structure has much the same form as the "observed" categorical structure: the $n \times n$ matrix $\mathbf{S}$ is decomposed into a binary $n \times m$ matrix $\mathbf{Z}$ (where m denotes the total number of categories, irrespective of system), and an $m \times m$ block diagonal "category correlation" matrix $\mathbf{W}$ (i.e., $\mathbf{S} = \mathbf{ZWZ}^\mathsf{T}$). In the latent feature model, each block of $\mathbf{W}$ is further decomposed into a set of binary features and weights: that is, we apply a second decomposition, also of the form $\mathbf{FVF}^\mathsf{T}$. Again, a very extensive literature exists providing methods for finding the best set of latent features (see Navarro, 2003).

Experiment

Method

Participants. Participants were 10 people recruited from the general community (4 males, 6 females). Ages ranged from 22 to 35 years.

Materials. The stimuli in this experiment were 20 greeble images, as shown in Figure 5. Ten males and ten females were selected, with four from each family and all with the same camera angle.

Procedure. The experiment began with a simple familiarization phase. Participants examined a sheet with all 20 greebles (with genders and families mixed up), in order to show participants the extent of variation in the set of greebles. Participants then completed the experiment individually, on a personal computer. For each trial, participants were presented with three greebles in a randomly chosen position on the screen. Participants first selected the "odd one out", then rated their confidence on a 5-point Likert scale ranging from 1 to 5, where 1 was "unsure" and 5 was "highly confident" (the confidence data is not discussed here). Each participant was

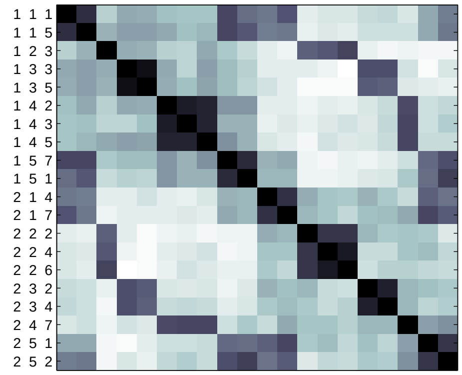

Figure 4: The empirical similarity matrix for 20 greebles obtained via the odd-one-out experiment. Darker tones indicate greater similarity. The labels adjacent to each row of the similarity matrix indicate which gender (first column) and which family (second column) the greeble belongs to, as well as the individual (third column).

presented with all possible 1,140 combinations of three greebles, in random order, divided up into 10 blocks. The instructions asked participants to work as "quickly and accurately" as possible. They were encouraged to take breaks between blocks. The experiment took around one hour to complete.

Results

Descriptive Analysis. Across the 11,400 judgments made in the experiment, each pair of greebles appeared between 180 and 187 times (the variation is due to small amounts of missing data). In a trial consisting of items i, j and k, if a participant chooses item k as the odd one out, we have evidence that items i and j are somewhat similar to one another. Accordingly, an empirical measure of similarity s_{ij} can be constructed as the proportion of trials in which items i and j are present, but neither is chosen as the odd one out (the standard error for these estimates at this sample size is approximately $\sigma \approx .04$). The data are shown in Figure 4: given the fact that there are non-trivial similarities between greebles that belong to different categories, and those similarities are not constant across categories (i.e., Figure 4 resembles Figure 3 more so than Figure 2) it appears that the similarity structure for the greebles is quite complicated, and probably not well-captured by a simple set of unrelated categories.

Model Selection. To test this more thoroughly, we fit a series of models to the data shown in Figure 4, using the Bayesian information criterion (BIC; Schwarz, 1978) to choose the best model for the data. BIC penalizes models for having excessive parameters that might result in overfitting the dataset and reducing the generalizability of the model (Myung & Pitt, 1997). The model that minimizes BIC is to be preferred. As illustrated in Table 1, when correlations between categories are disallowed (as per Eq. 1 and Figure 2), we are unable to capture human performance very well (BIC is large). If we allow the full range of possible associations, as is the case for the correlated categories model, performance improves substantially (a BIC difference of $1528 - 382 = 1146$ corresponds to an odds ratio of about $e^{1146/2} \approx 7 \times 10^{248}$ to 1 in favor of the full model). Of most interest, however, is the fact that this covariation is not

[2] The "constant" of proportionality that we use here and in Equation 4 is actually $\sqrt{w_{kk}^{(q)} w_{ll}^{(q)}}$, chosen by analogy to correlation coefficients. It *does* matter when fitting our models, but we suspect that it probably an unnecessary complication in the context of similarity modeling, and in future extensions we may drop this term entirely.

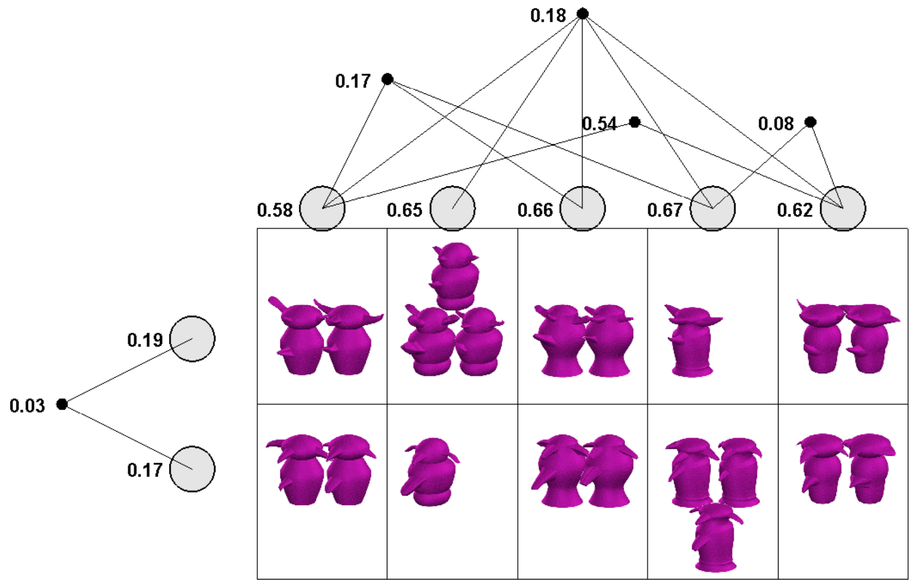

Figure 5: A latent feature model for the greebles choice data. The large grey circles adjacent to the table denote categories, with the two genders displayed as rows in the table, and the five families (1 to 5) shown as columns. The numbers displayed next to the grey circles are the category weights (i.e., diagonal elements of $\mathbf{W}$). Black dots represent inferred features: lines connect categories to the latent features that they possess, and the numbers show the weights associated with each feature.

Table 1: Performance of a subset of the models we considered, measured in terms of the percentage variance accounted for (VAF) and the Bayesian information criterion (BIC). The best model found is highlighted in bold. Models that do not allow correlated categories universally provide a poor account of people's decisions. The first two columns specify the structure used (uncorrelated, full-correlation, latent tree or latent features), while the last three describe the overall performance. Empty cells refer to cases where the corresponding categories were omitted entirely.

Gender	Family	VAF	# Param.	BIC
uncorrelated	-	14.4%	3	6130
-	*uncorrelated*	60.9%	6	2815
uncorrelated	*uncorrelated*	78.9%	8	1528
-	*uncorrelated*	82.3%	16	1602
uncorrelated	*full correlation*	95.8%	18	382
uncorrelated	*tree*	92.5%	13	566
uncorrelated	**feature**	**95.5%**	**12**	**356**

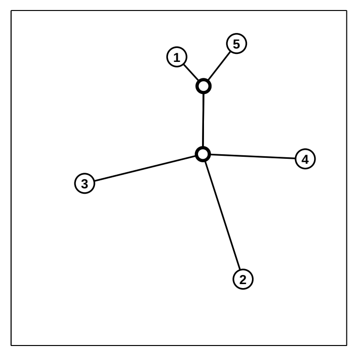

Figure 6: The topology of the optimal latent tree for the greeble families, explaining 92.5% of the variance in the data. Each leaf node (numbered circles) represents one of the five greeble families. Distance through the tree represents dissimilarity between families: in agreement with the latent feature model shown in Figure 5, the main regularity found is the similarities between families 1 and 5.

arbitrary: the latent feature model with four features shown in Figure 5 is superior (implied posterior odds ratio of about 442,413 to 1) to the full model. Qualitatively this is easy to see – the 6 extra parameters in the full model only explain an extra 0.3% of the variance relative to the latent feature model.

Nevertheless, it is important to note that not all kinds of latent structure work: in particular, the structure between families of greebles appears not to be hierarchical. Even the best tree we found (shown in Figure 6) was considerably worse than the full model, precisely because of the hierarchical constraint: as with with the latent feature model, the main regularity lies in the similarities between families 1 and 5. However, because the tree topology imposes strict hierarchical constraints, the less heavily weighted aspects to the latent feature model do not emerge within the tree. As a consequence, we were unable to find any category hierarchy that matched the performance of the simple correlated category model.

Interpretation. Model selection via the BIC incorporates a natural trade-off between data fit and model complexity. Since the preferred model relies on the assumption that there exists *structured* relationships between the different perceptual categories (via latent features), we can say with some confidence that this structure provides the best method for statistically representing the data set. However, to sustain a theoretical argument that the latent feature structure is the right psychological model, it should at least be the case that the extracted features be interpretable. This can certainly be done - the primary regularity (namely the similarities between families 1 and 5) extracted by both the featural and tree models) corresponds to something like "approximate convexity of body shape" (technically, the family 5 body shape is not convex, but comes very close). Families 2, 3 and 4 all involve major violations of convexity, but all in quite different fashions. Another latent feature corresponds to the distinction that families 4 and 5 are the only two "thin" families (families 1, 2 and 3 are all somewhat "curvy", but again in different fashions). The third latent feature, connecting families 1, 3 and 4, is much less convincing: the most plausible interpretation is that this feature should properly be assigned to family 5 as well, an expression of the fact that family 2 has a structurally different body shape (with two segments) to the other four families. That said, these are of course post-hoc descriptions and one should be careful in attaching much weight to them.

Discussion

While most research involving greebles has focused on their use as analogs for faces (e.g., Gauthier & Tarr, 1997), they are of some interest in their own right as moderately complicated visual stimuli. In fact, when considered in this light it seems surprising that no previous work has attempted to learn how people naturally perceive the greebles. Even in the context of the simple "odd one out" task we consider in this paper, it is clear that partitioning greebles into five families and two genders is an overly-simplistic way of representing the variation in body shapes that people automatically find. However, by providing explicit measures of similarity between greebles, it should be straightforward for future research to take these effects into account.

From a more theoretical perspective, by introducing explicit methods of uncovering the latent structure that underlies people's choices, it appears that these perceptual "categories" are not organized into a hierarchical organization, but rather seem to rely on a latent feature structure that can accommodate a richer pattern of variation. The idea of allowing correlations between features within the common features model that underlies techniques like additive clustering (Shepard & Arabie, 1979; Navarro & Griffiths, in press) is novel, but the restriction to this class of models is neither necessary nor desirable: in future, the approach could be extended to cover other representational models such as contrast models (Tversky, 1977; Navarro & Lee, 2004) or even hybrid models that mix between continuous and discrete properties (Navarro & Lee, 2003).

Acknowledgments. RGS was supported by an Australian Postgraduate Award, and DJN by an Australian Research Fellowship (ARC grant DP-0773794). We thank Nancy Briggs, John Dunn, Gert Storms and the reviewers for helpful comments, and Michael Tarr for making the greebles available. The similarity matrix presented in Figure 4 is available as a text file from DJN's website.

References

Behrmann, M., Marotta, J., Gauthier, I., Tarr, M. J., & McKeeff, T. J. (2005). Behavioral change and its neural correlates in visual agnosia after expertise training. *Journal of Cognitive Neuroscience*, *17*, 554-568.

Gauthier, I., & Tarr, M. J. (1997). Becoming a greeble expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673-1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nature*, *2*, 568-573.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. *Vision Research*, *38*, 2401-2428.

Griffiths, T. L., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (p. 381-386). Mahwah, NJ: Lawrence Erlbaum.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1433-1458.

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

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian aproach. *Psychonomic Bulletin and Review*, *4*(1), 79-95.

Navarro, D. J. (2003). *Representing stimulus similarity*. Unpublished doctoral dissertation, University of Adelaide.

Navarro, D. J., & Griffiths, T. L. (in press). Latent features in similarity judgment: A nonparametric Bayesian approach. *Neural Computation*.

Navarro, D. J., & Lee, M. D. (2002). Commonalities and distinctions in featural stimulus representations. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 685–690). Mahwah, NJ: Lawrence Erlbaum.

Navarro, D. J., & Lee, M. D. (2003). Combining dimensions and features in similarity-based representations. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems* (Vol. 15, pp. 67–74). Cambridge, MA: MIT Press.

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

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*(1), 54-65.

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

Rehder, B. (2006). When similarity and causality compete in category-based property generalization. *Memory & Cognition*, *34*, 3-16.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, *13*, 250-257.

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

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

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

Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York: Wiley.

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

Woodsworth, R. S., & Schlosberg, H. (1954). *Experimental psychology (3rd ed)*. London: Methuen.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Leaning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 585-593.