Psychonomic Bulletin & Review
2008, 15 (4), 732-749
doi: 10.3758/PBR.15.4.732

In search of abstraction: The varying abstraction model of categorization

Wolf Vanpaemel and Gert Storms
University of Leuven, Leuven, Belgium

A longstanding debate in the categorization literature concerns representational abstraction. Generally, when exemplar models, which assume no abstraction, have been contrasted with prototype models, which assume total abstraction, the former models have been found to be superior to the latter. Although these findings may rule out the idea that total abstraction takes place during category learning and instead suggest that no abstraction is involved, the idea of abstraction retains considerable intuitive appeal. In this article, we propose the varying abstraction model of categorization (VAM), which investigates the possibility that partial abstraction may play a role in category learning. We apply the VAM to four previously published data sets that have been used to argue that no abstraction is involved. Contrary to the previous findings, our results provide support for the idea that some form of partial abstraction can be used in people's category representations.

A classic question in cognitive psychology concerns what is stored as a consequence of learning a category, and hence what information people rely on when they make a categorization decision. It is generally assumed that learning a category involves the generation of a category representation and that assigning a novel object to a category involves the comparison of the object to that category representation. However, one of the most fundamental and unresolved issues in the categorization literature concerns the exact nature of this category representation.

Although the debate on category learning and category representation has a very long history, in the past few decades it has centered on the question of whether people represent a category in terms of an abstracted summary or a set of specific examples. Early work argued for the prototype view of category learning. Under this view, on the basis of experience with the category examples, people abstract out the central tendency of a category. In other words, a category representation consists of a summary of all of the examples of the category, called the prototype (see, e.g., Posner & Keele, 1968; Reed, 1972; Smith & Minda, 2002). The initial success of this view has gradually declined in favor of the exemplar view, in which experience with examples of a category does not lead to the development of an abstracted prototype; instead, people simply store all of the examples they encounter. In other words, a category representation consists of all of the individual examples of the category, called the exemplars (Brooks, 1978; Estes, 1986; Medin & Schaffer, 1978; Nosofsky, 1986).

The shift from the prototype to the exemplar view was motivated by several arguments. A first empirical argument for this shift involved the demonstration that exemplar models can account for empirical phenomena that were believed to provide evidence for the prototype view (e.g., the prototype enhancement effect; Busemeyer, Dewey, & Medin, 1984). A second empirical argument involved overwhelming evidence that exemplar models yield fits superior to those of prototype models in a wide variety of experimental settings (see Nosofsky, 1992, for a review).

The major argument against the prototype view, however, is that it fails to account for important aspects of human concept learning. In particular, a prototype does not seem to retain enough information about the examples encountered in learning. For example, prototypes discard information on correlations among features (e.g., large spoons tend to be made of wood, and small spoons are likely to be made of steel) and on the variability among the examples (e.g., U.S. quarters display very little variability in their diameters, whereas pizzas can vary greatly in size). This prototype feature is inconsistent with experimental studies that have suggested that people are sensitive to such information and store more about a category than just its central tendency (e.g., Fried & Holyoak, 1984; Medin, Altom, Edelson, & Freko, 1982; Rips, 1989).

As a consequence, the exemplar view is generally considered superior to the prototype view. However, the exemplar view also has not gone unchallenged. The main concern raised against this approach is its lack of any cognitive economy (Rosch, 1978). Under the exemplar view, people are assumed to store every training example and retrieve every exemplar from memory every time an item is classified. Both of these claims seem counterintuitive and excessive. For example, when people decide that a dog is a mammal, it seems unlikely that they compare the dog

W. Vanpaemel, wolf.vanpaemel@psy.kuleuven.be

Copyright 2008 Psychonomic Society, Inc.

to every single mammal they have ever encountered. The intuition that some cognitive economy is involved in category representations is confirmed by experimental findings suggesting that people store less information about a category than all of its members (Feldman, 2003).

In sum, the current theorizing on category representation involves a tension between informativeness and economy (Komatsu, 1992). A prototype representation has appealing economy but fails to provide the information people actually use. In contrast, an exemplar representation provides detailed information but is not economical. A natural way to resolve this tension would be to propose a representation that combines the benefits of both economy and informativeness. Such a representation would provide just enough representational information to describe the category structure in a sufficiently complete way.

Motivated by the appeal of such an intermediate representation, we propose the *varying abstraction model* (VAM; Vanpaemel, Storms, & Ons, 2005). It starts from the observation that the debate between the exemplar and prototype views can be usefully regarded as a debate on the use of *abstraction* in category representations. At the heart of the VAM is the idea that the category representations hypothesized by the exemplar and prototype views do not represent alternatives constituting a dichotomy, but rather correspond to the endpoints of a continuum: The exemplar representation corresponds to *minimal* abstraction, and the prototype representation corresponds to *maximal* abstraction. Between these endpoints, various new possible representations can be developed, balancing the opposing pressures of economy and informativeness. Such an intermediate representation would not consist of all exemplars, but neither would it consist of one single prototype. Instead, it would consist of a set of *subprototypes* formed by category members merging together. The intermediate representation would be less detailed and more economical than the exemplar representation, but more detailed and less economical than the prototype representation, corresponding to *partial* abstraction. On the basis of this extended class of representations, numerous categorization models can be developed, including the exemplar and prototype models. Crucially, all models of the VAM contrast only in their representational assumptions. Consequently, the VAM provides a simple framework for evaluating the idea that abstraction takes place during category learning.

The currently dominant practice when inferring the use of abstraction in category representations is to restrict the analysis to the extreme levels of abstraction—that is, to compare the prototype and exemplar representations only. This means that the wealth of intermediate representations corresponding to partial abstraction are overlooked. In light of the limitations of the extreme levels of abstraction, partial abstraction has considerable intuitive appeal. By formalizing the idea of partial abstraction, the VAM provides an alternative to the focus on exemplar and prototype representations only. It is important to highlight from the outset that the VAM is not intended as an improvement of the exemplar or of the prototype model. Rather, our intended contribution is to provide an improvement to the *debate* on abstraction in category representations, by providing a principled way to explore the use of abstraction in people's category representations.

Recently, a number of other authors have also proposed computational models that aim to go beyond the exemplar and prototype models. In particular, the rational model of categorization (RMC; Anderson, 1991), SUSTAIN (Love, Medin, & Gureckis, 2004), and the mixture models of categorization (MMC; Rosseel, 2002) share a starting point similar to that of the VAM. As will become clear in our General Discussion, the approach taken in the VAM differs in important ways from these earlier approaches. The main difference is that, unlike the other models, the VAM makes no strong assumptions about how representations arise, and therefore allows for a more general exploration of partial abstraction.

We organize our article by first reviewing briefly the best-known exemplar and prototype models. Next, we explain how the VAM positions these models as extremes on a continuum and formalizes models between the extremes. The VAM is then applied to four previously published data sets in order to evaluate the level of abstraction of people's category representations. Earlier analyses of these data sets failed to provide evidence in favor of abstraction. In contrast, the present VAM analysis shows that, for three of the four data sets, some form of abstraction took place during category learning. We also demonstrate that, for three data sets, our results are not caused by chance, because the different models encompassed by the VAM can be distinguished in a satisfactory way. Finally, we compare the VAM with related models and discuss some limitations and some possibilities for future research.

REVIEW OF THE EXEMPLAR AND PROTOTYPE MODELS

Both the exemplar and the prototype models assume that an object is classified as a member of a category if it is judged to be sufficiently similar to that category. The distinguishing assumption between the models is the exact nature of the category representation. Prototype models assume that a category is represented abstractly by the central tendency of the known category members (i.e., the prototype).[1] Categorization of an object depends on the relative similarity of the object to the prototypes of the relevant categories. By contrast, exemplar models assume that no abstraction is involved in category learning, but instead that a category is represented as the collection of its category members (i.e., the exemplars). Categorization of an object thus depends on the relative similarity of the object to all of the members of the relevant categories. In what follows, the formal descriptions of a widely tested exemplar model, the generalized context model (GCM; Nosofsky, 1986), and of its abstract counterpart, the multidimensional-scaling-based prototype model (MPM; Nosofsky, 1987; Reed, 1972), are reviewed.

Experimental Procedure

Category-learning tasks present people with stimuli and their accompanying category labels and require label

prediction for novel stimuli. A typical artificial category-learning task involves learning a two-category structure over a small number of stimuli. A subset of the stimuli are assigned to Categories A and B, and the remaining stimuli are left unassigned. Most experiments consist of a training (or category-learning) phase followed by a test phase. During the training phase, only the assigned stimuli are presented. The participant classifies each presented stimulus into either Category A or B and receives corrective feedback following each response. During the test phase, both the assigned and unassigned stimuli are presented. The unassigned stimuli are not seen in training, so they are novel to the participant. Because the assigned stimuli are used as the training stimuli, they are the basis for the category representation.

Stimulus Representation

Both the GCM and the MPM assume that stimuli are represented as points in a multidimensional psychological space. Such a multidimensional representation is typically derived from identification confusion data (see, e.g., Nosofsky, 1987) or from similarity ratings (see, e.g., Shin & Nosofsky, 1992) using multidimensional scaling (MDS; Borg & Groenen, 1997; Lee, 2001).

Once the stimuli are represented in a multidimensional space, the distances between the stimuli can be computed. There are several ways to compute the distance between a pair of stimuli (Ashby & Maddox, 1993). When $x_i = (x_{i1}, \ldots, x_{iD})$ denotes the coordinates of stimulus x_i in a D-dimensional space, the most common expression for the distance between the stimuli x_i and x_j is

$$d\left(x_i, x_j\right) = \left[\sum_{k=1}^{D} w_k \mid x_{ik} - x_{jk} \mid^r\right]^{1/r}. \quad (1)$$

Of crucial importance are the free parameters w_k, which model the psychological process of selective attention. The underlying motivation for this parameter is the assumption that when people are faced with a categorization task, they are inclined to focus on the dimensions that are relevant for the categorization task at hand and to ignore the ones that are irrelevant. In geometric terms, this mechanism of selective attention is represented in terms of stretching the space along the attended, relevant dimensions and shrinking the space along the unattended, irrelevant ones. As such, the parameters w_k can modify the structure of the psychological space. Since the parameters are constrained by $0 < w_k < 1$ and $\Sigma_{k=1}^{D} w_k = 1$, they can be interpreted as the proportion of attention allocated to dimension D_k and are often termed the *attention weights*. The differential weighting of dimensions has been a critical component of the GCM (and the MPM) and has enabled it to account for human categorization behavior.

The metric r is not a free parameter, but rather depends on the type of dimensions that compose the stimuli. In particular, previous investigations have supported the use of the city-block metric ($r = 1$) when stimuli vary on separable dimensions and the Euclidean metric ($r = 2$) when they vary on integral dimensions (see Shepard, 1991, for a review).

Stimulus-to-Stimulus Similarity

Both the GCM and the MPM assume that similarity is a decreasing function of distance in the psychological space, implying that similar stimuli lie close together, whereas dissimilar stimuli lie far apart (see, e.g., Nosofsky, 1984). In particular, the similarity between the stimuli x_i and x_j is given by

$$s(x_i, x_j) = e^{-cd(x_i,x_j)^{\alpha}}. \quad (2)$$

In this equation, c is a free parameter called the *sensitivity parameter*. It runs from 0 to ∞ and determines the rate at which similarity declines with distance. A high value of c implies that only stimuli that lie very close to each other are considered similar, whereas a low value of c implies that all stimuli are at least somewhat similar to each other.

Much as with the metric parameter r, the value of α depends on the nature of the stimuli and is not considered a free parameter. Two settings of the α parameter are prominent: $\alpha = 1$, resulting in an exponential decay function, and $\alpha = 2$, resulting in a Gaussian function. When the stimuli are readily discriminable, the exponential decay function seems to be the appropriate choice, whereas the Gaussian function is typically preferred when the stimuli are highly confusable (Shepard, 1987).

Stimulus-to-Category Similarity

Equation 2 can be used to compute the similarity of a stimulus to a certain category member. However, both models assume that a stimulus is classified according to its similarity to a *category*, not just to a category member. To go from stimulus-to-stimulus to stimulus-to-category similarity, a definition of a category is required. It is this assumption that distinguishes the GCM and the MPM.

In the GCM, a category is represented by all of its members, so the similarity of stimulus x_i to Category C_J is computed by summing the similarity of x_i to all of the category members of C_J:

$$\eta_{iJ} \equiv \sum_{x_j \in C_J} s\left(x_i, x_j\right). \quad (3)$$

In contrast, in the MPM, a category is represented by the category prototype, denoted as π_J. As such, the similarity of x_i to C_J equals the similarity of x_i to π_J:

$$\eta_{iJ} \equiv s(x_i, \pi_J). \quad (4)$$

Although the prototype generally does not match a stimulus, it is treated formally as a stimulus, thus $s(x_i, \pi_J)$ can be computed using Equations 2 and 1, given the coordinates of π_J. Since the prototype is the central tendency of all of the category members, the coordinates of π_J are simply the averaged coordinates of all of the n_J members of C_J:

$$\pi_{Jk} = \frac{1}{n_J} \sum_{x_j \in C_J} x_{jk}. \quad (5)$$

Response Rule

Both the GCM and the MPM assume that a categorization decision is governed by the Luce choice rule. Given

M relevant categories, the probability of categorizing stimulus x_i in Category C_J is then

$$p_{i,J} = \frac{\beta_J \eta_{i,J}}{\sum_{K=1}^{M} \beta_K \eta_{iK}}. \quad (6)$$

In this equation, every β_K is a free parameter, ranging from 0 to 1 and satisfying the constraint $\Sigma_{K=1}^{M} \beta_K = 1$. It is interpreted as the response bias toward Category C_K.

The response rule of Equation 6 is the one proposed in Nosofsky's (1986) original formalization of the GCM. Ashby and Maddox (1993) later generalized the response rule into

$$p_{i,J} = \frac{\beta_J \eta_{i,J}^{\gamma}}{\sum_{K=1}^{M} \beta_K \eta_{iK}^{\gamma}}. \quad (7)$$

This generalization involves the inclusion of an additional free parameter γ, termed the *response-scaling parameter*. It runs from 0 to ∞ and reflects the amount of determinism in responding. Values of γ larger than 1 reflect greater levels of determinism than are produced by Equation 6, and values of γ less than 1 reflect less determinism. Obviously, the generalized response rule reduces to the original response rule when $\gamma = 1$. The version of the GCM using this modified response rule is commonly referred to as *GCM-*γ.[2]

THE VAM

The GCM and the MPM are identical to each other in terms of their assumptions about stimulus representation, selective attention, similarity, and response rule. The assumption that distinguishes the two models is the category representation. Clearly, other representational possibilities can be hypothesized besides those considered by the GCM and the MPM. In this section, we develop a formal model of categorization that shares all of the common assumptions of the GCM and MPM but goes beyond these models in terms of the category representation.

Beyond the Exemplar and Prototype Representations

At the heart of the VAM is the idea that the exemplar and the prototype representations are the extreme endpoints on a continuum of abstraction. Along this continuum, positions between the extremes are held by representations in which an intermediate level of abstraction is assumed. Such an intermediate representation consists of a set of subprototypes formed by abstracting across a subset of category members. In particular, a category representation arises by partitioning[3] the category members into clusters and then averaging across the instances in each cluster. Crucially, in this way, the exemplar representation, the prototype representation, and a wealth of intermediate representations can be constructed. This is illustrated in Figure 1, which shows, for a category with five members represented in a two-dimensional space, the prototype representation (panel A), the exemplar representation (panel B), and an intermediate representation consisting of two subprototypes (panel C). Using this procedure of partitioning and averaging, a large set of representations can be created. The exhaustive set of possible representations for a category with four members is illustrated in Figure 2, represented in a two-dimensional space. The subprototypes representing the category are shown in black circles and are connected by lines to the original category members, shown in white circles.

The number of subprototypes in a representation can be interpreted as the *level of abstraction* of the representation, so that lower numbers of subprototypes correspond to more abstraction. As such, the 15 representations displayed in Figure 2 involve four different levels of abstraction. However, the representations do not only differ in their level of

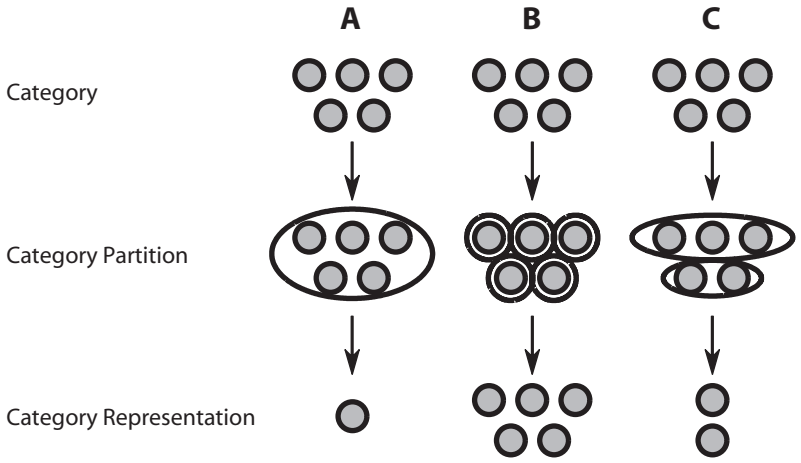

Figure 1. The two-step procedure to construct a category representation: (1) Partition the category into clusters and (2) construct the centroid for each cluster. In this way, it is possible to construct the prototype representation (panel A), the exemplar representation (panel B), and a set of intermediate representations, one of which is illustrated in panel C.

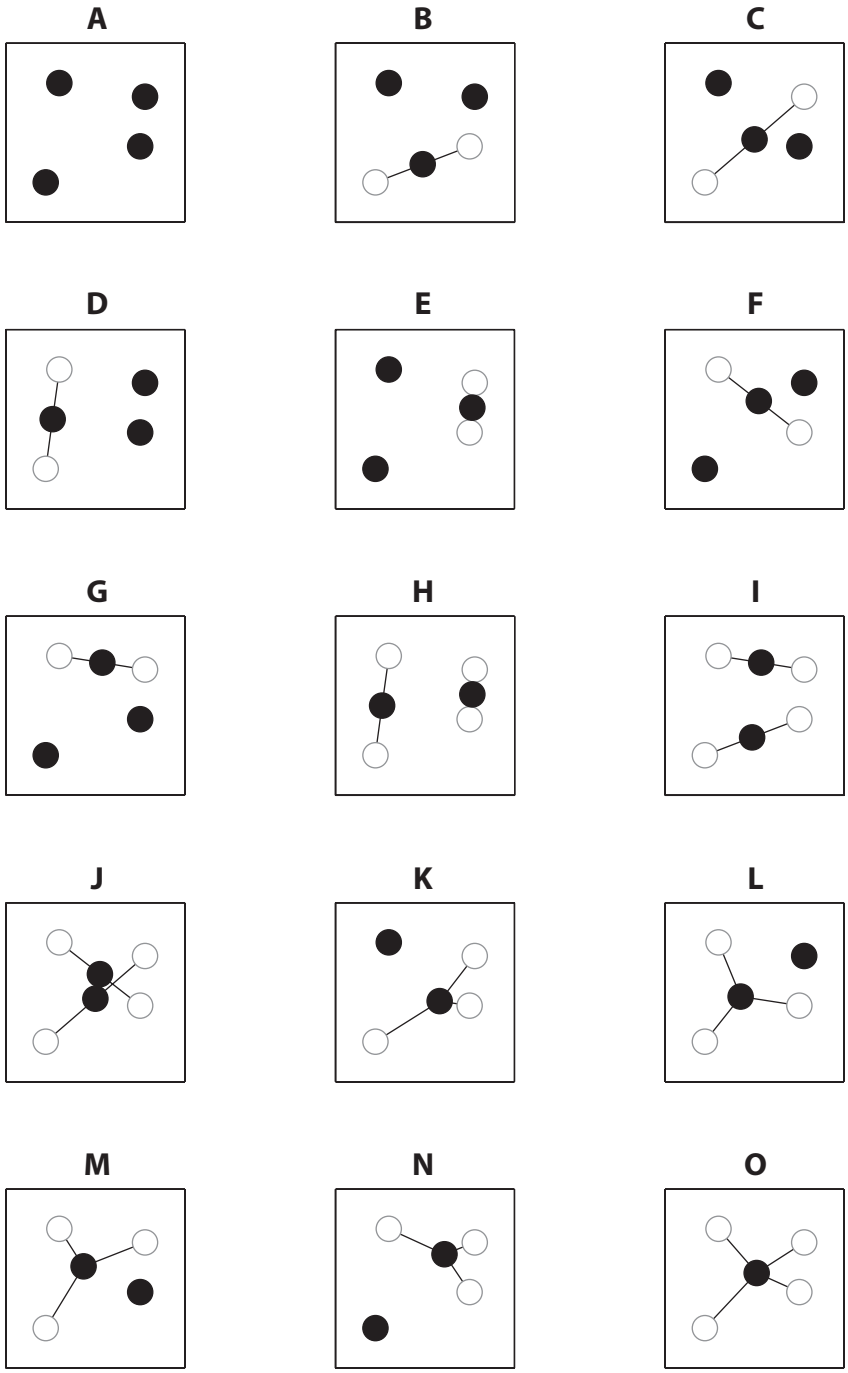

Figure 2. The 15 possible representations of the VAM for a category with four members in a two-dimensional space. The subprototypes are shown as black circles and are connected by lines to the original category members, shown as white circles. Panel A shows the exemplar representation, in which no category members are merged, and panel O shows the prototype representation, in which all four category members are merged in one single item. The remaining panels show all of the possible intermediate representations allowed by the VAM, in which a category is represented by three (panels B–G) or two (panels H–N) subprototypes.

abstraction. For all but the extreme levels of abstraction, the VAM proposes different representations at one single level of abstraction. In particular, in Figure 2, 6 of the representations are at a level of abstraction of three (panels B–G), and 7 are at a level of abstraction of two (panels H–N). Representations with the same level of abstraction share the number of subprototypes but differ in the category members that are merged. In particular, they can differ in the *degree of*

similarity that is involved in the merging of the category members. For example, in panel E, the category members being merged are much more similar to each other than the ones merged in panel C.

Formal Description of the VAM

Formally, a categorization model arises when, for every relevant category, a representation is combined with the assumptions shared by the GCM and the MPM. In particular, let $\mathbf{Q_J} = \{Q_1, Q_2, \ldots, Q_{q_J}\}$ denote a partition of Category C_J, and let n_i denote the number of category members in cluster Q_i. Further, let μ_i denote the centroid of Q_i, and $S_J = \{\mu_1, \mu_2, \ldots, \mu_{q_J}\}$ denote the set of all of the q_J centroids. These centroids are the subprototypes making up the category representation. The similarity of stimulus x_i to Category C_J is computed by summing the similarity of x_i to all q_J subprototypes representing C_J:

$$\eta_{iJ} \equiv \sum_{\mu_j \in S_J} s\left(x_i, \mu_j\right), \tag{8}$$

where $s(x_i, \mu_j)$ is the similarity of x_i to μ_j. Like the category prototype, the subprototypes can be treated formally as stimuli; thus, $s(x_i, \mu_j)$ can be computed by Equations 2 and 1, if the coordinates of the subprototypes are known. These are defined as the averaged coordinates of all the n_i category members within the cluster Q_i:

$$\mu_{ik} = \frac{1}{n_i} \sum_{x_j \in Q_i} x_{jk}. \tag{9}$$

Note that at the extreme values of q_J, the GCM and the MPM arise (i.e., Equations 3 and 4).

The VAM encompasses all of the categorization models that can be constructed by combining all of the possible representations of all of the relevant categories. In general, the number of possible partitions of a set of n elements is given by a number known as the nth Bell number, denoted by Bell(n). This implies that, in a categorization task with two Categories A and B (i.e., $M = 2$) with n_A and n_B category members, respectively, the VAM generates Bell(n_A) $\times$ Bell(n_B) different categorization models. A critical aspect of the VAM is that all models are matched to each other in every respect and contrast only in their representational assumptions. As such, all models have identical free parameters: $M-1$ response biases β_J (because of the constraint $\Sigma_{K=1}^{M} \beta_K = 1$), one sensitivity parameter c, one response-scaling parameter γ, and $D-1$ attention weights w_k (because of the constraint $\Sigma_{k=1}^{D} w_k = 1$), all summarized in the parameter vector $\boldsymbol{\theta} = (w_1, w_2, \ldots, w_{D-1}, c, \gamma, \beta_1, \beta_2, \ldots, \beta_{M-1})$. Balancing the models in terms of their parameters assures the fairest comparison between the different models.[4]

A VAM ANALYSIS OF EMPIRICAL DATA

Progress in understanding abstraction in category representation has often been sought by a systematic quantitative comparison of the GCM and the MPM. A review of these comparisons, as well as a theoretical treatment of the relations between prototype and exemplar models, is provided by Nosofsky (1992). His Table 8.5 summarizes fits of both the GCM and the MPM across 19 previously published data sets, involving a variety of category structures, experimental conditions, and types of stimuli. He concludes that "it is easily seen that the evidence is overwhelmingly in favor of exemplar-based category representations compared to prototype-based category representations, with the nature of the similarity rule held constant" (p. 163). Indeed, the MPM performed rather poorly relative to the GCM and provided a better fit than the GCM for only 1 data set. In sum, Nosofsky's (1992) review provides compelling evidence that a model that assumes no abstraction, like the GCM, generally fits empirical data better than a model that assumes total abstraction, like the MPM. Crucially, these findings rule out the use of total abstraction in category representations, but they cannot rule out the use of all forms of abstraction.

The studies reviewed by Nosofsky (1992) have been of considerable importance in the debate about the role of abstraction in categorization, so they seemed particularly attractive to be reanalyzed with the VAM. Therefore, in this section, the VAM is applied to 4 of the 19 data sets from Nosofsky's (1992) Table 8.5. The 4 data sets that were most appropriate for an initial VAM analysis were those with the smallest number of models implied by the design of the experiment, resulting in the selection of the data from Shin and Nosofsky's (1992) Experiment 3, Size 3, equal-frequency (E3S3EF) condition and from Nosofsky's (1987) saturation (A), saturation (B), and criss-cross conditions. All four conditions involved two categories to be learned, with a deterministic assignment of the category members to the categories.

In a VAM analysis of empirical categorization data, all models implied by the VAM are fit separately to the data. The most common method to fit a model to empirical data is maximum likelihood estimation (MLE; see, e.g., Myung, 2003). The idea behind MLE is to search for values of the free parameters that maximize the likelihood of observing the data.[5] The model yielding the smallest $-\ln L(\hat{\boldsymbol{\theta}})$ is selected as the best-fitting model.

In the VAM analysis of the four data sets, we tried to follow the original analyses as closely as possible. Obviously, the crucial difference between the analyses in the original studies and the VAM analysis was that, in the present study, the full set of representations as formalized by the VAM was considered, whereas in the original studies, generally only two representations were considered. Apart from this difference, the original analyses were followed in all major respects, in order to increase comparability. One minor difference between the present analysis and the original analyses was that we assumed nondifferential category bias (i.e., $\beta_K = 1/2$ for every K). The reason for this choice was that we wished to use as few free parameters as needed, and the analyses in both Shin and Nosofsky (1992) and Nosofsky (1987) revealed that response bias did not play a significant role. A second minor difference between the present analysis and the original analyses concerns the use of stimulus biases. Nosofsky (1987) fit-

ted a version of the GCM that made use of stimulus biases, which were estimated from the data of an identification experiment. Since this version of the GCM is not commonly used, we did not include the stimulus biases in the reanalysis of the data. In all other respects, we followed the original analyses: We used the categorization proportions from the original studies; we used the MDS solutions derived in the original studies; we assumed the Euclidean distance metric and the exponential decay similarity function (i.e., $r = 2$ and $\alpha = 1$), as in the original studies; and we did not include the response-scaling parameter γ in the analyses (i.e., $\gamma = 1$), as in the original studies.

The Shin and Nosofsky (1992) Data

The first set of data that we reanalyzed was from a series of experiments conducted by Shin and Nosofsky (1992) using the prototype-distortion paradigm (see, e.g., Posner & Keele, 1968). In this paradigm, generally, a category is defined by first creating a category prototype and then constructing the category members by randomly distorting these prototypes. Generalization is then tested by presenting the prototypes, the old distortions of them, and various new distortions.

Data and Results. In Experiment 3 of Shin and Nosofsky (1992), the stimuli used were random polygons. Two categories of polygons were created by first defining two prototypes and then generating 10 distortions of each one. In the Size 3 condition, for each category, 3 of these distortions were randomly selected as the training stimuli. An additional set of 5 transfer stimuli were created for each category, as follows. First, the prototype was redefined by calculating the average position of all 10 generated stimuli. Second, 2 new distortions were generated from each redefined prototype. Third, 2 new distortions were generated from a stimulus that was randomly selected from the training set.

For the main experimental manipulation, in a baseline condition all training stimuli were presented with the same frequency, whereas in a high-frequency condition, some of the training stimuli were presented more often than the others. However, because our main interest was in the category representation rather than in the effect of presentation frequency, we only analyzed the data from the baseline condition. In sum, we analyzed the data from the E3S3EF condition.

Thirty participants learned to classify the polygons into the two categories. Following a training phase in which feedback was provided after classification, a test phase was conducted during which all 6 training stimuli and all 10 transfer stimuli were presented. Classification feedback was provided only for the training stimuli. There were three blocks of test trials, with each stimulus presented once in each block, resulting in 90 classification decisions for every stimulus. The observed proportion of Category A responses for each stimulus, averaged across participants, is reported in Table 10 of Shin and Nosofsky (1992).

Following the test phase, all participants judged the degree of similarity between all pairs of the 16 polygons. Thirty other participants, who had not taken part in the categorization experiment, did the same. From the similarity judgment data, Shin and Nosofsky (1992) derived a four-dimensional MDS solution. This solution, reported in their Table A3, was taken as the underlying stimulus representation both for the theoretical analyses of Shin and Nosofsky and for the present VAM analysis.

In their theoretical analyses, Shin and Nosofsky (1992) found that the MPM fared poorly relative to the GCM, as reported in their Tables 11 and 13. In addition, they fitted a combined model, in which the relative contributions of the prototype and the exemplar representation could be estimated, and found that the parameter weighting the use of the prototype representation was 0 (see their Table 14). In sum, their analysis of the data in the E3S3EF condition did not provide any evidence for the operation of an abstraction process.

VAM analysis. Since the third Bell number (i.e., three members per category) is 5, the VAM encompasses 25 possible models. Table 1 shows the details of the VAM analysis for all the 25 models. Each model is described by two membership vectors, one for each category. In general, the representation of a category with n members using q subprototypes can be described in a convenient way by the membership vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, where $v_i \in \{1, 2, \ldots, q\}$ indexes the cluster membership of stimulus x_i. For example, for a category with five members, the exemplar representation is described by $\mathbf{v} = (1, 2, 3, 4, 5)$, the prototype representation is described by

Table 1
Summary Fits and Maximum Likelihood Parameters for All 25 Models Fitted to Shin and Nosofsky's (1992) E3S3EF Condition Data, As Ordered by Fit

Model		Fit		Parameters			
$\mathbf{v}_A$	$\mathbf{v}_B$	$-\ln L$	pvaf	w_1	w_2	w_3	c
1, 2, 3	1, 2, 3	63.45	91.57	0.48	0.02	0.11	1.50
1, 2, 2	1, 2, 2	66.38	90.99	0.32	0.11	0.27	1.66
1, 1, 2	1, 2, 2	67.91	89.87	0.43	0.04	0.19	1.61
1, 2, 2	1, 1, 2	68.96	89.82	0.40	0.10	0.00	1.88
1, 1, 2	1, 1, 2	71.82	89.05	0.42	0.22	0.02	2.06
1, 2, 1	1, 2, 2	72.62	89.03	0.52	0.04	0.17	1.47
1, 2, 3	1, 1, 2	76.52	88.74	0.46	0.15	0.00	1.79
1, 2, 1	1, 1, 2	78.99	87.52	0.58	0.13	0.03	1.63
1, 1, 2	1, 2, 3	79.58	85.29	0.46	0.02	0.25	1.55
1, 2, 2	1, 2, 3	80.08	85.52	0.55	0.01	0.35	1.35
1, 2, 1	1, 2, 3	80.15	85.28	0.54	0.00	0.22	1.50
1, 2, 3	1, 2, 2	85.10	85.00	0.42	0.14	0.11	1.59
1, 1, 1	1, 1, 1	85.30	85.17	0.54	0.02	0.03	1.69
1, 1, 2	1, 2, 1	89.93	83.25	0.63	0.05	0.02	1.46
1, 2, 1	1, 2, 1	90.48	83.35	0.73	0.01	0.02	1.41
1, 2, 2	1, 2, 1	92.68	82.89	0.93	0.03	0.00	1.11
1, 2, 3	1, 2, 1	100.28	79.47	0.85	0.03	0.00	1.27
1, 1, 1	1, 2, 2	110.67	75.83	0.50	0.04	0.38	1.36
1, 1, 1	1, 1, 2	110.94	77.86	0.67	0.03	0.30	1.22
1, 1, 2	1, 1, 1	112.04	76.10	0.44	0.12	0.02	2.00
1, 2, 1	1, 1, 1	112.38	78.11	0.37	0.13	0.01	2.30
1, 2, 2	1, 1, 1	118.80	74.58	0.48	0.18	0.00	1.70
1, 1, 1	1, 2, 1	160.09	63.40	0.67	0.04	0.09	1.54
1, 2, 3	1, 1, 1	178.33	57.47	0.45	0.21	0.00	1.88
1, 1, 1	1, 2, 3	180.19	56.85	0.46	0.00	0.36	1.67

Note—$\mathbf{v}_A$, $\mathbf{v}_B$, membership vector for Category A, B; $-\ln L$, negative value of the maximized log-likelihood; pvaf, percentage of variance accounted for; w_k, attention weight given to dimension D_k; c, sensitivity.

$\mathbf{v}$ = (1, 1, 1, 1, 1), and the intermediate representation of panel C in Figure 1 is described by $\mathbf{v}$ = (1, 1, 1, 2, 2).

In Table 1, the GCM and the MPM correspond to the models indexed by $\mathbf{v}_A = \mathbf{v}_B$ = (1, 2, 3) and $\mathbf{v}_A = \mathbf{v}_B$ = (1, 1, 1), respectively. The table reports, for every model, the negative log-likelihood and, as an auxiliary measure of fit, the percentage of variance accounted for. Although they are of lesser interest for the present goal, the best-fitting parameter values are reported as well. Our primary interest is in which representation best accounts for the observed data. In Table 1, the models are ordered by fit, with the best-fitting model on the top row. Apparently, of all 25 possible models considered, the model that captured the participants' performance best was the one that assumed an exemplar representation for both categories. Impressively, even when the level of abstraction was allowed to vary, there was no evidence for the presence of abstraction in the category representations.

The Nosofsky (1987) Data

Apart from the prototype-distortion paradigm, a second influential research paradigm involves simple perceptual stimuli that vary along a few salient dimensions. The remaining three data sets we reanalyzed were collected by Nosofsky (1987) using this paradigm.

Data and Results. Nosofsky (1987) conducted a color categorization study using a stimulus set of 12 Munsell color chips varying in brightness and saturation. On the basis of this set of stimuli, six different category structures were constructed, as shown in Figure 5 of Nosofsky (1987). The three category structures studied in the present article are the saturation (A), saturation (B), and criss-cross structures. Figure 3 illustrates these category structures in the psychological space. To derive these representations, Nosofsky (1987) instructed 34 participants to identify all 12 stimuli, and he used the data from this identification experiment to derive the MDS solution, reported in his Table 3.

Twenty-four other participants learned to categorize the same set of 12 stimuli in both the saturation (A) and criss-cross conditions, and 40 others were assigned to the saturation (B) condition. In the saturation (A) condition, participants were presented with one block of 120 trials, and in the two other conditions, two blocks of 120 trials were presented. Each stimulus was presented 10 times in each block. After classifying a stimulus in either of the two categories, feedback was given only in the case in which a stimulus was assigned to a category. Table 4 of Nosofsky (1987) shows the proportion of Category A responses for each stimulus in each condition, averaged across participants. The saturation (A) data were obtained during the final 90 trials of the single block, and those for the two other conditions were obtained during the second block, resulting in sample sizes of 180, 400, and 240 for the saturation (A), saturation (B), and criss-cross conditions, respectively.

As shown in Nosofsky's (1987) Tables 5 and 6, he found that, for the saturation (A) condition, the MPM yielded essentially the same fit as the GCM. For the saturation (B) condition, the MPM was found to fit substantially worse

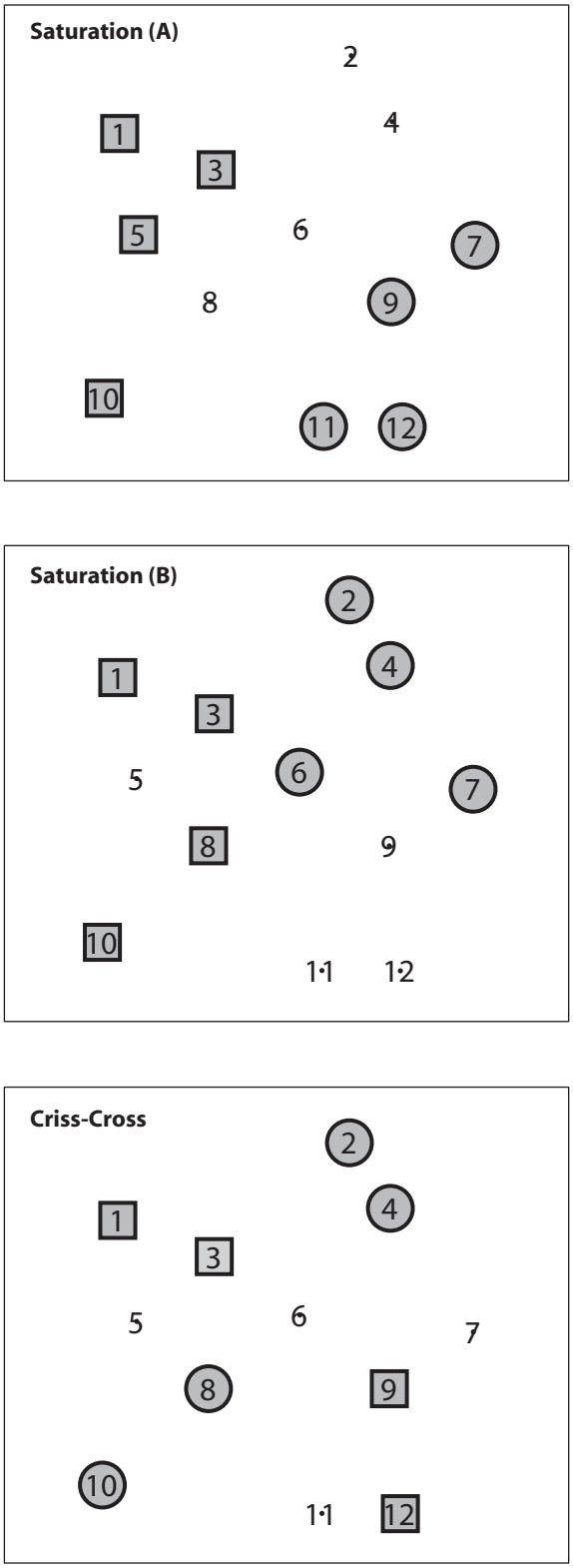

Figure 3. Schematic representation of Nosofsky's (1987) saturation (A), saturation (B), and criss-cross conditions in the psychological space. Squares denote training stimuli assigned to Category A, and circles denote those assigned to Category B. The remaining stimuli are unassigned. Adapted from Nosofsky (1987).

than the GCM. Finally, the MPM was unable to account for the data in the criss-cross condition, whereas the GCM yielded an impressive fit. In sum, once again no evidence for the operation of an abstraction process could be discerned on the basis of these data.

Especially in the criss-cross condition, the evidence against such a process was particularly compelling. It is instructive to understand why the MPM failed so dramatically in this condition. Inspection of the category structure explains why the prototype representation was insufficient as a basis for categorization: Apparently, the centroids for Categories A and B virtually overlap. If the representation for Category A were identical to that for Category B, the similarity of a stimulus to Category A would be identical to its similarity to Category B, and the model would predict performance at chance. Therefore, it is far from surprising that the MPM failed to account for the data collected in the criss-cross condition.

Importantly, for the criss-cross structure, several intermediate representations seem highly plausible. In this structure, a category can be split up in two subcategories,[6] so it is reasonable to expect that, when abstraction takes place in this condition, subprototypes would be based on the subcategories. This observation led Nosofsky (1987) to test one multiple-prototype representation in the criss-cross condition, consisting of four subprototypes. It is shown in Figure 4, using the graphical conventions adopted earlier (i.e., the subprototypes are shown in black and are connected by lines to the original category members, shown in white). Nosofsky (1987) found that, although this multiple-prototype model fared far better than the MPM, the GCM was still superior. Thus, no evidence for the use of abstraction could be provided once again.

VAM analysis. Clearly, other multiple-prototype representations than the one considered by Nosofsky (1987) can be formalized. This is exactly what the VAM provides. All three conditions involved two categories of four members each, implying 225 possible models. Table 2 shows the results of the VAM analysis. The table reports, for each condition, the negative log-likelihood, the percentage of variance accounted for, and the parameters for the best-fitting VAM instantiation.

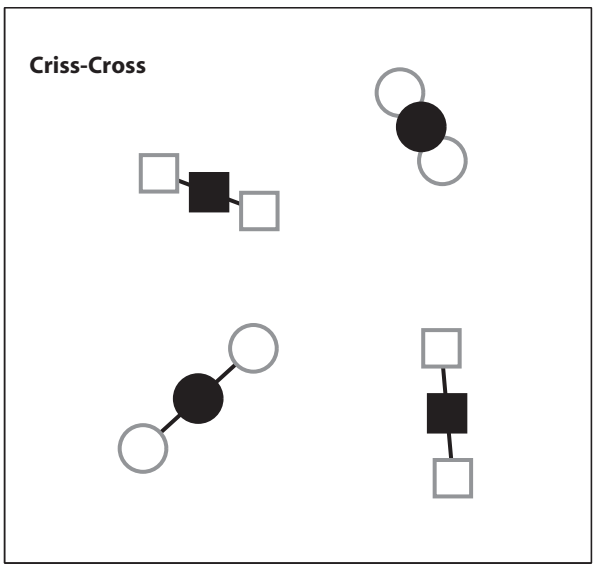

Figure 4. The intermediate representation tested by Nosofsky (1987) in the criss-cross condition.

Table 2
Summary Fits and Maximum Likelihood Parameters for the Best-Fitting Model to Nosofsky's (1987) Saturation (A), Saturation (B), and Criss-Cross Data

	Fit		Parameters	
Condition	$-\ln L$	pvaf	w_1	c
Saturation (A)	41.09	98.52	0.79	1.06
Saturation (B)	56.45	99.04	0.75	1.23
Criss-cross	44.65	99.09	0.62	1.60

Note—$-\ln L$, negative value of the maximized log-likelihood; pvaf, percentage of variance accounted for; w_k, attention weight given to dimension D_k; c, sensitivity.

Whereas the focus of the present research lies on the representation, we briefly mention that the estimated values of the free parameters in the best-fitting models are intuitively acceptable. Particularly in the two saturation conditions, in which the first dimension was clearly more diagnostic than the other dimension for performing the categorization task, the estimated attention weights on the first dimension—.79 and .75 for the saturation (A) and (B) conditions, respectively—are consistent with the expectation that the participants attempted to attend selectively to the first dimension. Furthermore, the weights are in close correspondence with the estimated weights for the GCM and MPM in the original study.

Our primary interest, however, is which of the representations describes the observed data best. Figure 5 shows the best-fitting representation for each condition. Inspection of these figures reveals that when the level of abstraction was allowed to vary in each condition, the representation yielding the best fit assumed some form of partial abstraction. In the saturation (B) condition, Category A adopted the exemplar representation, but in all other best-fitting representations at least two category members were merged. A comparison of these VAM results with those produced by the GCM is provided in the Appendix.

The best-fitting representation in the saturation (A) condition seems somewhat counterintuitive, since in Category A rather disparate category members are being merged. The representation gains some psychological plausibility in the space modified by selective attention, as is illustrated in the top panel of Figure 5B. In contrast, the best-fitting representation in the saturation (B) and criss-cross conditions are psychologically easily interpretable. In the saturation (B) condition, Category A adopts the exemplar representation, and in the Category B representation, two rather similar category members are merged. In the criss-cross condition, the best-fitting representation has a particularly strong intuitive appeal. As already hypothesized by Nosofsky (1987), this representation

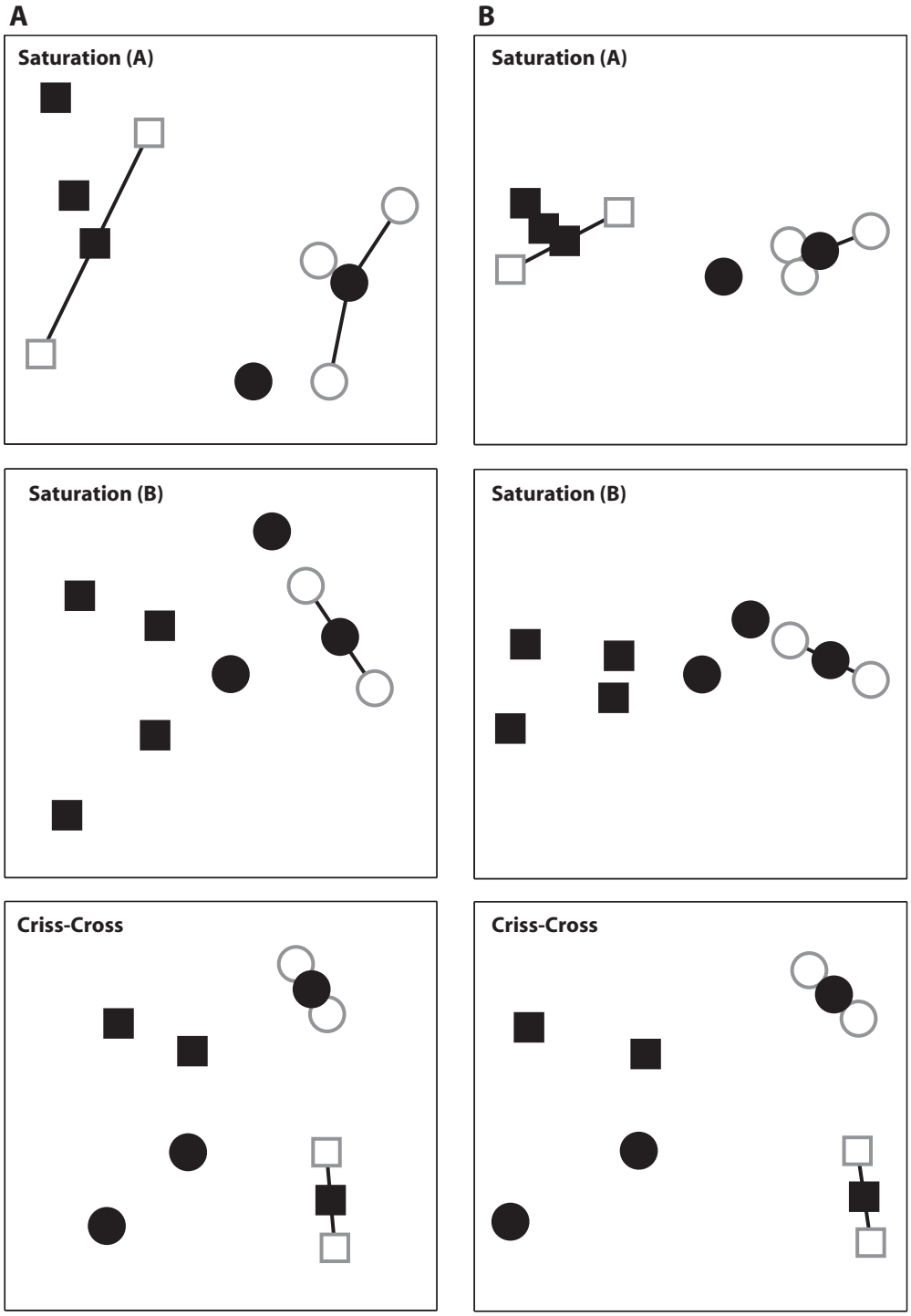

Figure 5. The best-fitting representations for Nosofsky's (1987) saturation (A), saturation (B), and criss-cross conditions, in a psychological space either without (column A) or with (column B) modification by selective attention.

involves the formation of subprototypes for the subcategories. However, unlike the intermediate representation considered by Nosofsky (1987), shown in Figure 4, the intermediate representation providing the best fit to the data does not consist of four subprototypes, but of only two. It is formed by merging two stimulus pairs, 2–4 and 9–12, but unlike the representation in Figure 4, Stimuli 1, 3, 8, and 10 are left unmerged. In sum, in contrast to the earlier conclusions, the results of the VAM analysis of Nosofsky's (1987) empirical data provide support for the idea that some form of abstraction is involved in people's category representations.

THE DISTINGUISHABILITY OF THE REPRESENTATIONS

Since the VAM enlarges the set of representational possibilities, a potential concern with the varying abstraction approach is that it considers too many representations. In particular, the problem is that, for any data set, there will always be a representation that yields a better fit than the others, but evaluating whether the superior fit of this representation is reliable or accidental is difficult. Therefore, a question of central importance is whether the representations are distinguishable. Obviously, if the discrimination between the different representations fails, the results obtained in the previous section would be due to chance, and the conclusions we reached not legitimate.

To gain information regarding the distinguishability of the representations, for all four conditions we conducted a large-scale recovery simulation study. Such a study involves artificial data for which the true underlying representation is known but that are contaminated by sampling variability. Of interest is the ability of the VAM to recover the true, data-generating representation when it is fit to the simulated data. If the VAM analysis is not governed by chance, it should be able to "see through" the random variation caused by sampling error and correctly discern the representation that generated the data.

In the remainder of this section, we first provide details on the procedure used in the recovery study and then report its results for both the Shin and Nosofsky (1992) and the Nosofsky (1987) data.

Procedure

The data were generated from a particular model by the following three steps: selecting a set of parameter values, computing the classification probabilities according to the model, and adding sampling error to these probabilities. Since we wished to generate response patterns that were similar to those observed, the parameter values were obtained by fitting the generating model to the empirical data set. Given the optimal parameter values for this model, a classification probability of a stimulus according to the model was then obtained by substituting these parameter values in Equation 7. Sampling error was introduced by generating a set of binary-valued responses (0 or 1) from the binomial probability distribution corresponding to the classification probability. In each application, the number of binary responses was equal to the sample size of the empirical study, again in order to generate response patterns similar to those observed (Pitt, Kim, & Myung, 2003). Once this procedure had been applied for each of the stimuli, an artificial data set as generated from the model was obtained. From each model in the VAM, we generated 100 such artificial data sets.

Each of these artificial data sets was treated exactly as an empirical data set, so each model of the VAM was fit to each of them in order to determine the best-fitting model. As a result, each artificial data set was classified to a certain model. Since the VAM was fit to every artificial data set, this implies that $25 \times 100 \times 25 = 62{,}500$ models were fit for the E3S3EF condition and, likewise, $225 \times 100 \times 225 = 5{,}062{,}500$ models were fit for each condition from Nosofsky (1987). For all four applications together, over 15 million models were fit.

The Shin and Nosofsky (1992) Data

Table 3 shows, for each of the 25 different models, the recovery rate r_i and the false recovery rate f_i. The recovery rate of model M_i ($i = 1, 2, \ldots, 25$) is the percentage of *correctly* classified data sets *from* M_i. It is defined by $r_i = \text{ncorr}_i/\text{ngen}_i$, where ngen_i denotes the number of generated data sets from model M_i (i.e., 100 for every i, in the present case) and ncorr_i denotes the number of correctly classified data sets generated by M_i (i.e., the number of data sets for which M_i both generated the data and was selected as the best-fitting model). The false recovery rate of model M_i is the percentage of cases in which data sets were *incorrectly* classified *to* M_i. It is defined by $f_i = \text{nfalse}_i/\text{nfalse}$, where nfalse denotes the total number of incorrectly classified data sets across all models and nfalse_i denotes the number of data sets incorrectly classified to M_i (i.e., the number of data sets for which M_i did not generate the data but was nonetheless selected as the best-fitting model).

Globally, the recovery is quite good, with the individual recovery rates ranging from 49% to 100%. Of all 2,500 artificial data sets, 280 were incorrectly classified. As indicated by the false recovery rates shown in the last column, the model responsible for most of the misclassifications was the GCM. However, this happened for only 30 [i.e., $(10.71 \times 280)/100$] data sets. This seems negligible, since as many as 2,400 data sets had been generated by

Table 3
Recovery Rates and False Recovery Rates for All 25 Models of Shin and Nosofsky's (1992) E3S3EF Condition

$\mathbf{v}_A$	$\mathbf{v}_B$	r_i	f_i
1, 1, 1	1, 1, 1	72.00	4.29
1, 1, 1	1, 1, 2	95.00	3.21
1, 1, 1	1, 2, 1	97.00	0.36
1, 1, 1	1, 2, 2	92.00	1.07
1, 1, 1	1, 2, 3	100.00	0.36
1, 1, 2	1, 1, 1	89.00	0.00
1, 1, 2	1, 1, 2	95.00	6.07
1, 1, 2	1, 2, 1	77.00	9.29
1, 1, 2	1, 2, 2	84.00	6.43
1, 1, 2	1, 2, 3	82.00	9.64
1, 2, 1	1, 1, 1	99.00	2.50
1, 2, 1	1, 1, 2	90.00	6.43
1, 2, 1	1, 2, 1	78.00	8.57
1, 2, 1	1, 2, 2	85.00	6.43
1, 2, 1	1, 2, 3	79.00	7.14
1, 2, 2	1, 1, 1	97.00	0.36
1, 2, 2	1, 1, 2	98.00	3.93
1, 2, 2	1, 2, 1	49.00	2.86
1, 2, 2	1, 2, 2	92.00	4.29
1, 2, 2	1, 2, 3	89.00	2.14
1, 2, 3	1, 1, 1	99.00	0.00
1, 2, 3	1, 1, 2	99.00	1.43
1, 2, 3	1, 2, 1	98.00	0.36
1, 2, 3	1, 2, 2	96.00	2.14
1, 2, 3	1, 2, 3	89.00	10.71

Note—$\mathbf{v}_A$, $\mathbf{v}_B$, membership vector for Category A, B; r_i, recovery rate for model M_i (in %); f_i, false recovery rate for model M_i (in %).

models other than the GCM. In sum, it seems that in the experimental design (which includes the category structure, the number and positions of the transfer stimuli, and the sample size) of the E3S3EF condition, the different VAM representations are sufficiently distinguishable to legitimize the conclusions of the VAM analysis in the previous section.

The Nosofsky (1987) Data

Each of the three conditions of the Nosofsky (1987) study involved 225 models, so a detailed report of the recovery rates for each model separately would take up too much space. Instead, the 225 individual recovery rates are depicted graphically in Figure 6 using a histogram. Furthermore, to get a global picture of the recovery in these three conditions, the second column of Table 4 reports the overall recovery rates r. The overall recovery rate averages the individual recovery rates r_i (i.e., $r = \Sigma_{i=1}^{225} \text{ncorr}_i/22{,}500$) and corresponds to the percentage of correctly classified data sets across all models for each of the conditions.

In the saturation (A) condition, recovery was clearly poor, with the correct model recovered for less than 50% of the data sets. Figure 6 shows that some of the models had a very high recovery rate and others a very low recovery rate, but that for the bulk of the models recovery was poor. In the saturation (B) condition, the overall recovery rate was reasonably high, with the true model recovered about 9 times out of 10. As shown in Figure 6, for some exceptional models, recovery was moderate or even poor, but for most of the models recovery ranged from very good to perfect. Impressively, in the criss-cross condition, recovery was virtually perfect: As few as 265 of the 22,500 artificial data sets were incorrectly classified, and the vast majority of the models were perfectly recovered (i.e., 203 of the 225 models had an individual recovery rate of 100%).

Table 4 also reports the recovery and false recovery rates for three models with a privileged status. The models deserving special attention are the MPM, the GCM, and the model that fitted the empirical data best (see Figure 5). The most interesting finding is that, in all three conditions, the recovery rate for the MPM was particularly bad. In fact, in the saturation (B) condition, the MPM had the worst recovery of any of the models, and in the saturation (A) and criss-cross conditions, only two models had worse recovery.[7] In contrast, the GCM had perfect recovery in two of the conditions. Only in the saturation (A) condition was the GCM not clearly distinguishable.

Considering the false recovery results, none of the three models displayed in Table 4 was responsible for a large amount of the misclassifications in the saturation (A) and (B) conditions. In fact, in these conditions, no model at all had a particularly large false recovery rate; the largest were 0.78% and 2.24%, respectively. In the criss-cross condition, the false recovery rate was also very small for the GCM and, most importantly, for the best-fitting model. Quite surprisingly, the MPM was responsible for more than 6% of the misclassifications. In total, 6 models were responsible for at least 5% of the misclassifications each. However, since the number of misclassified data sets was

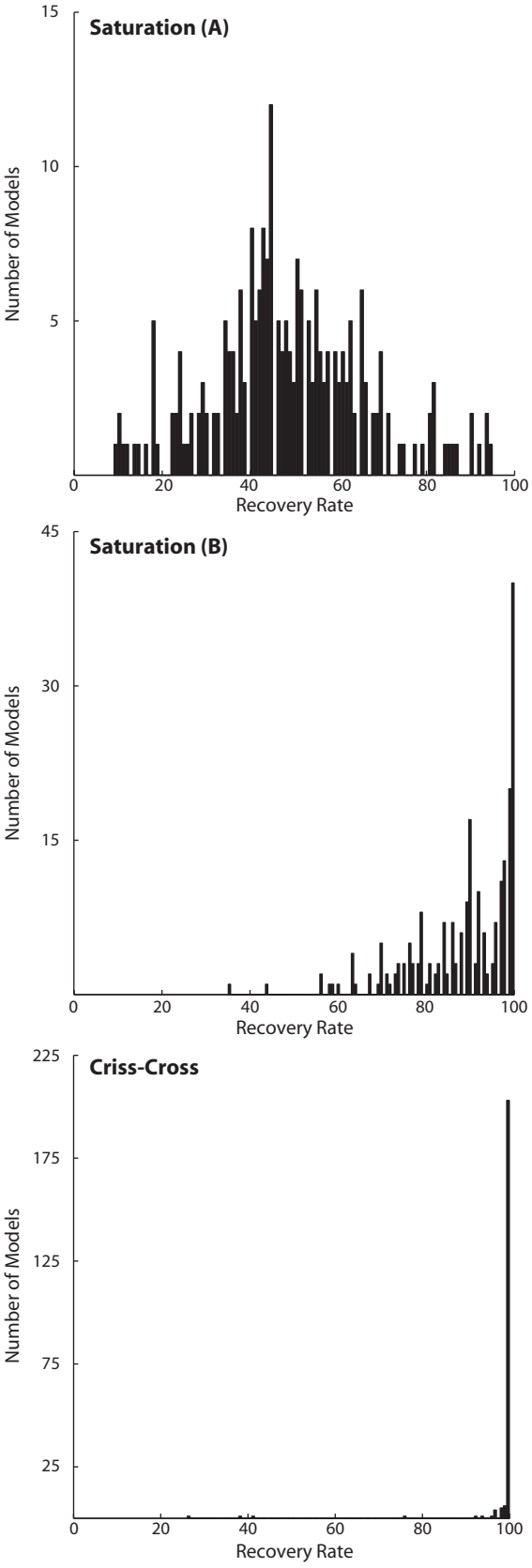

Figure 6. Histogram of the recovery rates for all models separately in the saturation (A), saturation (B), and criss-cross conditions.

Table 4
Recovery Statistics for Nosofsky's (1987) Saturation (A), Saturation (B), and Criss-Cross Conditions

Condition	r	r_{MPM}	r_{GCM}	r_{best}	f_{MPM}	f_{GCM}	f_{best}
Saturation (A)	48.50	10.00	18.00	61.00	0.50	0.54	0.40
Saturation (B)	88.68	35.00	100.00	100.00	1.61	0.04	0.08
Criss-cross	98.82	41.00	100.00	100.00	6.42	0.38	0.00

Note—r, overall recovery rate (in %); r_M, recovery rate for model M (in %); f_M, false recovery rate for model M (in %).

so low, even the model with the highest false recovery rate (9.43%) misclassified only 25 of the 22,400 data sets that were not generated by the model itself (0.11%).

In sum, the recovery simulation study indicated that the experimental design of the saturation (A) condition was not appropriate for distinguishing the different representations. In the other two conditions, however, recovery ranged from very good [the saturation (B) condition] to excellent (the criss-cross condition). Apparently, the experimental design of these conditions allowed for discriminating the different representations.

GENERAL DISCUSSION

In this article, we proposed the VAM, which is designed to help us understand what people store as a result of learning a category. Much of the previous research has contrasted models assuming minimal abstraction with models assuming maximal abstraction. However, the intriguing possibility that learning a category results in a representation at an intermediate level of abstraction has been largely overlooked. Such a representation has considerable intuitive appeal, because it balances the opposing pressures of cognitive economy and informativeness. The VAM greatly expands the set of representational possibilities traditionally considered, by incorporating representations at a partial level of abstraction. The prototype and exemplar representations are special, extreme cases of the VAM. Between these extremes, a wide variety of previously ignored representations are formalized.

The VAM provides a convenient framework for evaluating different representations, since all models of the VAM are identical to each other in every respect except the representation they assume. As such, when behavioral data are analyzed using the VAM, the goal is to evaluate whether abstraction is involved in category learning when allowance is made for partial abstraction, rather than to test the VAM per se. With this goal in mind, four previously published data sets were reanalyzed using the VAM. Earlier analyses of these data sets did not support the presence of abstraction in category learning.

The results of a recovery simulation study implied that in Nosofsky's (1987) saturation (A) condition, the different representations of the VAM were indistinguishable, and therefore that the fit of the representation that captured those data best was not reliably better than the fit of the other representations. Interestingly, of the best-fitting representations in the four conditions we reanalyzed, that selected in the saturation (A) condition was the most counterintuitive one, since very dissimilar category members were merged.

Furthermore, the simulations demonstrated that in Shin and Nosofsky's (1992) E3S3EF condition and Nosofsky's (1987) saturation (B) condition, the different VAM representations were fairly distinguishable, so the conclusions of the VAM analysis seem legitimized in these conditions. For the E3S3EF data, no evidence for abstraction could be provided, but the representation that captured the saturation (B) data best did assume some abstraction.

Finally, in Nosofsky's (1987) criss-cross condition, the representations were almost perfectly distinguishable. In this condition, both categories were made up of subcategories, which made partially abstract representations particularly appealing. The representation revealed by the VAM analysis here was easily interpretable psychologically, since the subprototypes were based on the subcategories. Given the near-perfect distinguishability of the representations, the a priori appeal of an intermediate representation, and the high plausibility of the best-fitting representation, we believe that the evidence for abstraction in the criss-cross condition is very compelling. In sum, contrary to the dominant view that no abstraction is involved in category representations, the present results make a strong case that abstraction can be involved in category representations.

Comparison With Related Models

The VAM is certainly not the first model to go beyond the prototype and exemplar representations. A first attempt to combine the benefits of both representations was to balance them (see, e.g., Busemeyer et al., 1984; Homa, Dunbar, & Nohre, 1991; Medin, Altom, & Murphy, 1984; Shin & Nosofsky, 1992). Basically, the authors of these studies included a mixture parameter that weighted the relative contributions of an exemplar and a prototype representation. The VAM is clearly different from such a hybrid approach. A hybrid model balances the two standard category representations, whereas the VAM explores a set of new category representations.

A number of other proposed models are much closer to the VAM. In particular, the RMC (Anderson, 1991; recently extended by Griffiths, Canini, Sanborn, & Navarro, 2007), SUSTAIN (Love et al., 2004), and the MMC (Rosseel, 2002) assume, along with the VAM, that categories are represented by multiple subprototypes formed by category members merging together. In the remainder of this section, we briefly summarize these models and indicate how our approach differs from theirs.

In the RMC (Anderson, 1991; Griffiths et al., 2007), clusters of individual stimuli are created during the learning process. These clusters are formed incrementally and are used to represent the categories. The probability that a training stimulus is assigned to a cluster is a function both of the similarity of that stimulus to the cluster's central tendency and of the prior probability of each cluster. The prior probability is partly determined by a free parameter, known as the *coupling parameter*, that influences the degree to which stimuli are clustered together during the learning process. A small value of the coupling parameter increases the probability that a stimulus will be assigned to a new cluster, whereas a large value of the parameter will cause stimuli to join other stimuli in preexisting clusters. The probability of assigning a stimulus to a certain category depends on the similarity of that stimulus to the central tendency of each cluster.

In SUSTAIN (Love et al., 2004), categories are also represented by one or more clusters of stimuli. Again, the cluster representation grows incrementally. It starts simple, with one cluster centered on the stimulus presented first. The representation then grows by assigning a stimulus to the cluster to which it is most similar. SUSTAIN is inherently biased toward a low number of clusters, in the sense that a new cluster is recruited only in response to a surprising event (e.g., a prediction error in supervised learning, or encountering a stimulus that is not similar enough to any existing cluster in unsupervised learning). Categorization is then based on the activation of these clusters. SUSTAIN includes attention weights that can be adjusted by learning rules to focus selectively on dimensions that are diagnostic in the categorization task. Also the association weights, linking the clusters to the categories, are adjusted by learning rules.

Finally, the MMC (Rosseel, 2002) provides a very general framework in which numerous different models of classification can be expressed. In particular, it is possible to define large families of semiparametric classifiers in which categories are represented by a finite mixture distribution. One specific version of the MMC, called the *reduced exemplar model* (REX), shares many commonalities with the VAM. In this model, a category is represented by a number of exemplar nodes. Only stimuli that are sufficiently different from an exemplar node are stored separately in memory, whereas a stimulus that is similar to an existing exemplar node will contribute to that node.

The obvious commonality between the RMC, SUSTAIN, the REX, and the VAM is that a category representation is based on the merging of category members. Despite this commonality, however, there are also key differences between the VAM and the models discussed above.

A first dimension of contrast concerns the basis of the clustering. The RMC, SUSTAIN, and the REX all assume that clustering is driven by similarity. By contrast, the VAM makes no a priori assumptions about which members should be merged together or kept separate. Consider, for example, panels C and E of Figure 2. The representation of panel C involves the merging of two rather dissimilar category members, whereas in the representation of panel E two very similar category members are merged. Both representations are generated by the VAM, whereas a model in which clustering is driven by similarity would only consider the representation of panel E.

The second dimension of contrast concerns the bias toward the number of clusters. In the RMC, this bias can be set using the coupling parameter. In SUSTAIN, the representation is initially directed to a low number of clusters (which is referred to as a *simplicity bias*). In the MMC, there is also a bias toward a low number of clusters, although it is rather vague (e.g., "For model selection, we simply choose the model containing the smallest number of mixture components, but still fitting the data reasonably well"; Rosseel, 2002, p. 192). By contrast, the VAM has no built-in preference for either a low or a high number of clusters, but considers all possible levels of abstraction.

It would be fairly straightforward to extend the VAM with a mechanism in which similar category members are merged and that is biased toward abstract representations (i.e., a low number of clusters). However, we deliberately chose not to. Similarity-based clustering and a simplicity bias are intuitively plausible, but we preferred to remain agnostic about how a representation is generated. The obvious advantage of this approach is that, by also incorporating representations in which dissimilar category members are merged and representations that are sufficiently detailed, the VAM allows for a more general investigation of the idea of partial abstraction than the related, more constrained approaches do.

Limitations

Going beyond the prototype and exemplar representations in an attempt to balance economy and informativeness represents an important research approach. Nevertheless, our study had some limitations.

Modeling averaged data. A legitimate concern is that, in all four applications, the VAM was fit to data that were averaged over participants. Model fits involving averaged data can be misleading, because such data can obscure patterns at the level of the individual participant. For example, in the VAM analysis of the averaged data, we implicitly assumed that all participants relied on the same representation, which is clearly a simplifying assumption. However, the main goal of the present research was to reanalyze data that have provided support in favor of the view that no abstraction takes place during category learning, in order to evaluate the potential presence of partial abstraction. Therefore, we kept our analysis as close as possible to the original analysis, to allow comparisons between the original results and ours. Since average classification accuracies were reported and analyzed in the original studies, we believe that in the present study, modeling averaged data was appropriate.

The response-scaling parameter. A second concern that can be raised against the present VAM reanalysis is that we used the original response rule of Equation 6 instead of the generalized response rule of Equation 7. The inclusion of the response-scaling parameter would allow our model to potentially account for responses ei-

ther more or less deterministic than those produced by the probability-matching response rule of Equation 6. The main reason for fixing $\gamma = 1$ was that, as mentioned above, we tried to hold constant all major aspects of the modeling in the original studies, except for the representational possibilities. Because in the original studies the GCM was used rather than the GCM-γ, the inclusion of γ would have blurred the comparison with the original studies. A second reason was that we wished to use as few free parameters as necessary. We do not expect that including γ would have had a large impact on the modeling results, since it has been shown that the inclusion of the γ parameter is most appropriate when individual data from highly experienced participants are modeled (see, e.g., McKinley & Nosofsky, 1995), and in this article only averaged data were analyzed.

The number of representations. As discussed above, in comparison with related approaches, the VAM is fairly unconstrained. Keeping the VAM unconstrained has the advantage that it allows for a general exploration of partial abstraction, but it comes at a price. The number of representations encompassed by the VAM grows quickly with the number of category members.[8] Therefore, the main limitation of the VAM is that, without introducing constraints on which representations should be considered, it can only be applied to data collected in experiments with small categories. Fortunately, experiments with a small number of training stimuli have a strong tradition in studies with artificial categories, so there is a wealth of data to which the VAM can be applied. For example, one of the most commonly used structures is the 5–4 structure, introduced by Medin and Schaffer (1978). In this design, one category has 5 members and the other category has 4, which implies that the VAM generates 780 models. This is sufficiently low to make a VAM analysis computationally feasible.

One way of dealing with the problem of the large number of representations encompassed by the VAM for large categories would be to consider only a single representation at each level of abstraction. As such, a category with n members would generate n representations only, rather than Bell(n). To select the single representation at a given level of abstraction, a reasonable strategy would be to adopt a similarity-based clustering procedure, comparable to the clustering implemented in the RMC, SUSTAIN, and REX. In particular, the selection of the representation could be based on any of a range of similarity-based cluster algorithms—for instance, average-link clustering, K-means clustering (Hastie, Tibshirani, & Friedman, 2001), or hierarchical clustering (Johnson, 1967). An application of a K-means variant of the VAM to natural language concepts can be found in Verbeemen, Vanpaemel, Pattyn, Storms, and Verguts (2007). However, as mentioned above, similarity-based selection of the representations might result in good-fitting intermediate representations being missed, and therefore would limit the exploration of partial abstraction. The selection would exclude not only representations that seem very sensible, but others as well that may seem implausible a priori but that could be formed as a result of presentation order, presentation frequency, or category contrast effects, or as a result of specific instructions given during the experimental task.

Directions for Future Research

Despite these limitations, the idea that people make use of an abstract category representation appears to be viable and to deserve continued investigation. An important avenue that needs to be pursued is to provide more converging evidence regarding the use of abstraction. Therefore, more data sets need to be analyzed, both on the group level and the individual-participant level.

Providing more evidence for the use of abstraction is only a preliminary research goal. The most challenging future goal will be to understand the conditions under which abstraction could be expected to play a role in category learning. In other words, instead of focusing on the question of *whether* abstraction occurs, the most exciting focus will be on *when* it occurs. Surprisingly little is known about the conditions in which people are likely to form an abstract representation of a category. Recent work has begun to suggest variables that facilitate the construction of an abstracted representation. For example, Minda and Smith (2001) found the prototype representation to be superior to the exemplar representation when categories are large. Thus, the level of abstraction may relate to the size of the category. Second, Feldman (2003) suggested that the exemplar representation is appropriate for a highly complex category, whereas a simple category is better represented by a prototype. Thus, the level of abstraction may be influenced by the complexity of the category. Finally, Smith and Minda (1998) found that, for large categories, neither the exemplar nor the prototype representation was superior at all points in the learning trajectory. Instead, a prototype representation dominated early in the learning process but slowly gave way to an exemplar representation. Thus, the level of abstraction may be related to the exact point in the learning process.

In sum, these studies suggest that human conceptual structure is sufficiently flexible to adopt highly abstracted representations at times, but that it can also accommodate highly differentiated representations at others (but see, e.g., Nosofsky & Zaki, 2002, for a contrasting view). People may adapt the level of abstraction of their category representations according to category size and complexity, or over time. Most previous studies have focused exclusively on the exemplar and prototype representations, and therefore have never been in a position to carefully investigate the influence of these variables on the level of abstraction. In contrast, the VAM provides a convenient framework for tracing out in considerable detail what factors influence the level of abstraction in people's category representations. Therefore, we believe that the VAM will advance progress in our understanding of the nature of category representations.

AUTHOR NOTE

The research reported in this article is part of an interdisciplinary project sponsored by the Research Council of the University of Leuven (Grant IDO/02/004, given to the second author). This research benefited

greatly from suggestions by Brad Love, Lance Rips, Ed Wisniewski, Safa Zaki, and several anonymous reviewers. We are particularly grateful to Dan Navarro, for his many inspiring comments, and Michael Lee, for his invaluable help and encouragement. Correspondence concerning this article should be addressed to W. Vanpaemel, University of Leuven, Department of Psychology, Concat, Tiensestraat 102, B-3000 Leuven, Belgium (e-mail: wolf.vanpaemel@psy.kuleuven.be; Web site: ppw.kuleuven.be/concat).

REFERENCES

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.

Balasubramanian, V. (1997). Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions. *Neural Computation*, **9**, 349-368.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169-211). Hillsdale, NJ: Erlbaum.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 638-648.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, **18**, 500-549.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, **12**, 227-232.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 234-257.

Griffiths, T. L., Canini, K. R., Sanborn, A. N., & Navarro, D. J. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 323-328). Mahwah, NJ: Erlbaum.

Grünwald, P. (2000). Model selection based on minimum description length. *Journal of Mathematical Psychology*, **44**, 133-152.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Homa, D., Dunbar, S., & Nohre, L. (1991). Instance frequency, categorization, and the modulating effect of experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 444-458.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241-254.

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

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, **45**, 149-166.

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

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, **28**, 41-50.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 128-148.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 37-50.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 333-352.

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

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 775-799.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, **47**, 90-100.

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

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87-108.

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

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 924-940.

Pitt, M. A., Kim, W., & Myung, I. J. (2003). Flexibility versus generalizability in model selection. *Psychonomic Bulletin & Review*, **10**, 29-43.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, **109**, 472-491.

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

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.

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

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, **46**, 178-210.

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

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 53-71). Washington, DC: American Psychological Association.

Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, **121**, 278-304.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 800-811.

Vanpaemel, W., Storms, G., & Ons, B. (2005). A varying abstraction model for categorization. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 2277-2282). Mahwah, NJ: Erlbaum.

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

NOTES

1. Other conceptions of a prototype have been proposed. For example, a prototype could be an idealized caricature of all of the cat-

egory members (see, e.g., Lynch, Coley, & Medin, 2000). Arguably, the idea of a prototype as the central tendency of a category is the most widespread.

2. It has been noted (e.g., by Ashby & Maddox, 1993) that the response-scaling parameter γ is not identifiable in the MPM, so the generalized response rule cannot be incorporated in that model.

3. A partition of a set S is a collection of disjoint, nonempty subsets (or clusters) of S whose union is S.

4. As mentioned in note 2, γ cannot be incorporated in the MPM, since it is not identifiable. Importantly, in all of the other models defined by the VAM, including the GCM, γ is identifiable.

5. The set of parameter values maximizing the likelihood is denoted by $\hat{\boldsymbol{\theta}}$. For computational efficiency, instead of maximizing the likelihood, generally the natural logarithm of the likelihood is maximized. This transformation does not affect the estimated value of $\hat{\boldsymbol{\theta}}$. Of interest is the negative value of the maximized log-likelihood, $-\ln L(\hat{\boldsymbol{\theta}})$, for which a smaller value indicates a better fit. When an experimental design involves two categories, as is the case for all of the applications considered in this study, the binomial likelihood is assumed, so the log-likelihood is given by

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{m} \Big\{ k_i \ln p_i(\boldsymbol{\theta}) + (t_i - k_i) \ln\big[1 - p_i(\boldsymbol{\theta})\big] + \ln t_i! - \ln k_i! - \ln(t_i - k_i)! \Big\}. \tag{10}$$

In this equation, k_i denotes the number of times stimulus x_i ($i = 1, \ldots, m$) was judged to belong to Category A out of the t_i trials in which it was presented, and $p_i(\boldsymbol{\theta})$ represents the probability of classifying stimulus x_i in Category A, as defined by Equation 7.

6. According to the labeling scheme of the stimuli in Figure 3, these subcategories are formed by the stimulus pairs 1–3, 9–12, 2–4, and 8–10.

7. Note that in the E3S3EF condition, as well, there was only one model with worse recovery than the MPM (see Table 3).

8. For instance, the first 10 values of the Bell number are 1, 2, 5, 15, 52, 203, 877, 4,140, 21,147, and 115,975.

APPENDIX
Comparing the GCM and the VAM

As we have stressed from the outset, the VAM is a framework to evaluate the level of abstraction of the category representations that people use. The question of interest in a VAM analysis is not whether the VAM is superior to the GCM as a cognitive model, but rather which representation captures the data best. In other words, despite being based on the GCM, the VAM is not intended, and should not be regarded, as a competitor for the GCM. In fact, applying the VAM can provide support for the GCM, as was the case in our reanalysis of the E3S3EF data. However, if warranted, a quantitative comparison of the GCM and the VAM is possible. Since in the E3S3EF condition the GCM was the best-fitting VAM instantiation, there is not much point in comparing the two models for those data, since parsimony would select the least complex model (i.e., the GCM). Therefore, we restrict this discussion to the three Nosofsky (1987) data sets. For these three sets, the VAM provides a better fit to the data than the GCM.

A first means of comparison is provided by Table A1. It shows, for each of the three conditions, the negative log-likelihood, the percentage of variance accounted for, and the parameters for the GCM; it can be used as a source for comparison with Table 2. Furthermore, Figure A1 provides a scatterplot of the observed against the model-based Category A response proportions for the three conditions, separately for the GCM and the best-fitting VAM instantiation, as identified in Figure 5.

Table A1
Summary Fits and Maximum Likelihood Parameters for the GCM to Nosofsky's (1987) Saturation (A), Saturation (B), and Criss-Cross Data

	Fit		Parameters	
Condition	$-\ln L$	pvaf	w_1	c
Saturation (A)	46.05	97.53	0.73	1.01
Saturation (B)	58.82	98.69	0.68	1.18
Criss-cross	50.36	98.66	0.34	1.44

Note—Above, $-\ln L$ is the negative value of the maximized log-likelihood; pvaf, the percentage of variance accounted for; w_k, the attention weight given to dimension D_k; c, the sensitivity.

When comparing the VAM and GCM, however, the main question of interest is whether the VAM as a whole, rather than one particular VAM instantiation, provides a better fit to the data than the GCM. Since the VAM includes the GCM as a special case, the VAM will always fit any data set at least as well as the GCM. Consequently, what really matters is not whether the VAM fits the data better than the GCM, but rather whether the improvement in fit provided by the VAM is significant. To address this issue, minimum description length (MDL; see, e.g., Grünwald, 2000) can be used.[A1] The MDL value of a model can be expressed by a sum of terms, one of which relates to the fit of a model, the rest of which reflect the model complexity (see, e.g., Pitt, Myung, & Zhang, 2002). In particular, the MDL value balances the lack of fit and the complexity, so a smaller value is better. Balasubramanian (1997) suggested that the complexity term for MDL is approximately a measure of the logarithm of a count of the number of "distinguishably" different probability distributions indexed by a model, so the challenge is to count the number of distinguishably different probability distributions within the VAM. Although daunting in principle, performing this count is easy if one assumes that all models of the VAM are

APPENDIX (Continued)

equally complex. Under this assumption, if the VAM implies n models in a particular experimental setting, the VAM is exactly n times more complex than any single VAM instantiation, such as the GCM. This implies that the difference in MDL values for the VAM and the GCM would be approximately

$$\begin{aligned}\Delta\text{MDL} &\equiv \text{MDL}_{\text{VAM}} - \text{MDL}_{\text{GCM}} \\ &\approx -\ln L(\hat{\boldsymbol{\theta}}_{\text{VAM}}) + \ln L(\hat{\boldsymbol{\theta}}_{\text{GCM}}) + \ln \text{Bell}(n_{\text{A}}) + \ln \text{Bell}(n_{\text{B}}). \end{aligned} \quad \text{(A1)}$$

Since a smaller MDL value is better, a negative value of ΔMDL would indicate that the VAM is preferred over the GCM, and a positive value would indicate that the GCM is the superior model.

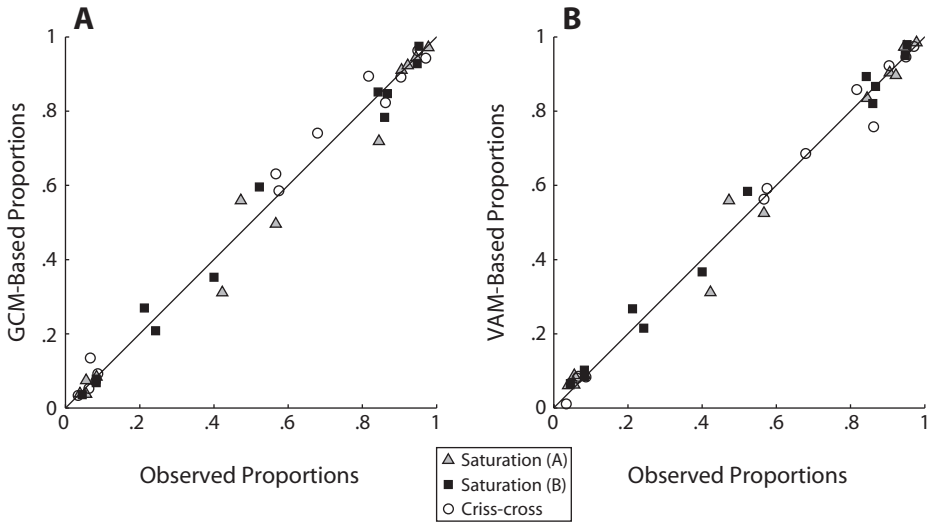

Figure A1. Scatterplot of the observed against the model-based Category A response proportions in the three conditions from Nosofsky (1987) for both the GCM (panel A) and the best-fitting VAM instantiation (panel B).

The approximation above rests on the assumption that all models of the VAM are equally complex. Pitt, Kim, and Myung (2003) noted that a recovery simulation study, such as the one reported in this article, may serve as a crude measure for the relative complexities of a set of models. The logic is that, if a model comparison method that ignores model complexity performs reasonably well, the differences in complexity among the models involved are small. Since the model comparison method used in the present recovery study (i.e., MLE) is insensitive to model complexity but performed reasonably well in the saturation (B) and criss-cross conditions, the assumption that all models were of roughly equal complexity is probably valid for these two conditions. In the saturation (A) condition, recovery was much worse, so we cannot conclude that the models there were equally complex. Therefore, the approximation of Equation A1 is legitimized only in the saturation (B) and criss-cross conditions.

A comparison of the log-likelihoods displayed in Tables 2 and A1 indicates that the differences in fit between the GCM and the VAM were 2.37 and 5.71 in the saturation (B) and criss-cross conditions, respectively. In both cases, the two categories had four members each (i.e., $n_{\text{A}} = n_{\text{B}} = 4$). Since the fourth Bell number is 15, the penalty complexity term equals $2 \times \ln 15 = 5.42$. This penalty is larger than the difference in fit in the saturation (B) condition, but smaller than the difference in the criss-cross condition. Hence, when comparing the VAM as a whole to the GCM, in the saturation (B) condition the GCM should be preferred, but in the criss-cross condition the VAM is superior to the GCM, even when the VAM is penalized for its complexity.

A1. This suggestion was made by Dan Navarro.

(Manuscript received August 24, 2006;
revision accepted for publication December 7, 2007.)