

2 The generalized context model: an exemplar model of classification

Robert M. Nosofsky

Model description

Conceptual overview

According to the *generalized context model* (GCM) (Nosofsky, 1986), people represent categories by storing individual exemplars (or examples) in memory, and classify objects based on their similarity to these stored exemplars. For example, the model assumes that people represent the category of ‘birds’ by storing in memory the vast collection of different sparrows, robins, eagles, ostriches (and so forth) that they have experienced. If an object is sufficiently similar to some of these bird exemplars, then the person would tend to classify the object as a ‘bird’. This exemplar view of categorization contrasts dramatically with major alternative approaches that assume that people form abstract summary representations of categories, such as rules or idealized prototypes.

The standard version of the GCM adopts a *multidimensional scaling* (MDS) approach to modelling similarity relations among exemplars (Shepard, 1958, 1987). In this approach, exemplars are represented as points in a multidimensional psychological space. Similarity between exemplars is a decreasing function of their distance in the space. In many applications, a first step in the modelling is to conduct similarity-scaling studies to derive MDS solutions for the exemplars and to discover their locations in the multidimensional similarity space (Nosofsky, 1992b).

A crucial assumption in the modelling, however, is that similarity is not an invariant relation, but a highly context-dependent one. To take an example from Medin and Schaffer (1978), humans and mannequins may be judged as highly similar in a context that emphasizes structural appearance, but would be judged as highly dissimilar in a context that emphasizes vitality. In the GCM, the context-dependent nature of

The writing of this chapter was supported by grants FA9550-08-1-0486 from the Air Force Office of Scientific Research and MH48494 from the National Institute of Mental Health.

similarity is modelled in terms of a set of selective-attention weights that systematically modify the structure of the psychological space in which the exemplars are embedded (Carroll & Wish, 1974). As will be illustrated below, the weights serve to ‘stretch’ the psychological space along highly attended, relevant dimensions, and to ‘shrink’ the space along unattended irrelevant dimensions. This stretching and shrinking can have profound influences on similarity relations among exemplars and on the resulting classification predictions from the model.

Finally, the model assumes that the individual exemplars may be stored in memory with differing memory strengths. The memory strength of an exemplar may be influenced by factors such as frequency of presentation, recency of presentation, different forms of feedback provided during learning, and so forth. When a test item is presented so as to be classified, the exemplars that are most likely to be retrieved (and therefore to influence most strongly the classification decision) are those that are highly similar to the test item and that have high memory strengths. However, because exemplar retrieval is a probabilistic process, all exemplars stored in memory may influence classification decision making.

The conceptual ideas summarized above are illustrated schematically in Figure 2.1. Consider first the top panel. We suppose that there are two categories, A and B, with five exemplars in each category. The exemplars are composed of two dimensions. Exemplars A2 and B4 are close together in the space, so are highly similar to one another; whereas exemplars A5 and B2 are far apart, so are highly dissimilar. (The symbols used to illustrate the exemplars vary in their size in order to illustrate that the exemplars may be stored in memory with differing strengths.) Imagine that the observer is asked to classify test stimulus i into one of the two categories. According to the model, the observer sums the similarity of test item i to all of the exemplars of Category A and to all of the exemplars of Category B (weighted by the memory strengths of the exemplars), and the classification decision is based on the relative magnitude of these summed similarities. In the top panel of Figure 2.1, test item i is roughly equally similar to the exemplars of the two categories, so the observer would classify the test item into the two categories with roughly equal probability.

Notice, however, that the horizontal dimension is far more relevant than is the vertical dimension for discriminating between the members of the two categories. (That is, all exemplars of Category B tend to have large values along the horizontal dimension, whereas all exemplars of Category A tend to have small values along the horizontal dimension.) Presumably, an experienced observer would learn this aspect of the

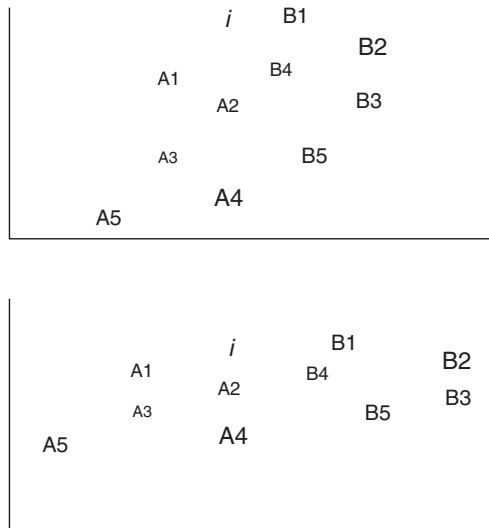


Figure 2.1 Schematic illustration of a category structure to explain the workings of the GCM. The top panel illustrates the category structure with equal attention to both dimensions. The bottom panel illustrates the structure following selective attention to the horizontal dimension.

category structure, so would learn to give greater attention to the more relevant horizontal dimension than to the less relevant vertical dimension. The bottom panel of Figure 2.1 illustrates the same category structure, but now assuming the above-described selective-attention strategy. The space is ‘stretched’ along the more-attended horizontal dimension and is ‘shrunk’ along the less-attended vertical dimension. In effect, by implementing this selective-attention strategy, the observer is attempting to optimize similarity relations for the given classification task (Nosofsky, 1984, 1986). There is now greater separation between the exemplars of contrasting categories (lowered between-category similarity), yet less separation among the exemplars within each category (heightened within-category similarity). Furthermore, note that this selective-attention strategy has a profound influence on the classification predictions from the model. In the top panel, item *i* was roughly equally similar to the exemplars of the two categories; however, following selective attention to the relevant dimension (bottom panel), the test item is now far more similar to the exemplars of Category A. Thus, in the latter situation, the exemplar model predicts that the test item would be classified into Category A with high probability.

Computational assumptions

In this section I present a brief description of how the GCM is formalized. The description assumes that there is an initial training phase in which observers are presented with n unique training exemplars. The training phase is followed by a test phase in which both training items and new transfer items might be presented. On each trial during the test phase, the observer classifies the test item into one of K_N categories. According to the model, the probability with which item i is classified into Category \mathcal{J} during the test phase is given by:

$$P(\mathcal{J} | i) = \frac{b_{\mathcal{J}} \left[\sum_{j=1}^n V_{j\mathcal{J}} s_{ij} \right]^{\gamma}}{\sum_{K=1}^{K_N} b_K \left[\sum_{k=1}^n V_{kK} s_{ik} \right]^{\gamma}}, \quad (1)$$

where s_{ij} denotes the similarity between item i and exemplar j ; $V_{j\mathcal{J}}$ denotes the memory strength of exemplar j with respect to Category \mathcal{J} ; γ is a freely estimated response-scaling parameter ($0 < \gamma$); and $b_{\mathcal{J}}$ ($0 < b_{\mathcal{J}}$) denotes the response-bias for Category \mathcal{J} . Thus, according to the model, the observer sums the similarity of item i to all exemplars j belonging to Category \mathcal{J} , weighted by their Category \mathcal{J} memory strengths (and by any differential response bias). This summed similarity constitutes the ‘evidence’ in favour of Category \mathcal{J} . This evidence is then divided by the summed evidence for all of the categories to predict the Category- \mathcal{J} classification probability.

The parameter γ in Equation 1 is a response-scaling parameter that influences the degree of determinism in classification responding (Ashby & Maddox, 1993; Nosofsky & Zaki, 2002). When $\gamma = 1$, the observer responds probabilistically by ‘probability matching’ to the relative summed similarities of each category; whereas when γ grows greater than 1, observers respond more deterministically with the category that yields the largest summed similarity. The memory-strength values ($V_{j\mathcal{J}}$) in Equation 1 are typically not free parameters but rather are given by the nature of the experimental design. Usually, they are set equal to the relative frequency with which each exemplar j is provided with Category \mathcal{J} feedback during the classification training phase. (In the most usual classification learning paradigms, the exemplars are presented with equal frequency and each exemplar is assigned to only a single category. In that simple case, all exemplars j that are assigned to Category \mathcal{J} receive

memory strengths equal to 1; whereas the memory strength of an exemplar with respect to all its unassigned categories is equal to 0.)

To apply Equation 1, one needs to compute the similarity between item i and each exemplar j , s_{ij} . In the standard version of the GCM, each exemplar j is represented as a single point in an M -dimensional psychological space. Let x_{jm} denote the value of exemplar j on Dimension m . The distance between item i and exemplar j is given by the weighted Minkowski power model,

$$d_{ij} = \left[\sum_{m=1}^M w_m |x_{im} - x_{jm}|^r \right]^{1/r}, \quad (2)$$

where the value r determines the form of the distance metric. In situations involving highly separable-dimension stimuli (Garner, 1974; Shepard, 1964), the value r is typically set equal to 1, which yields a city-block distance metric. By contrast, in situations involving integral-dimension stimuli, the value r is set equal to 2, which yields a Euclidean distance metric. The w_m values in Equation 2 are freely estimated attention-weight parameters (with $0 \leq w_m \leq 1$, and $\sum w_m = 1$), reflecting the degree of attention that observers give to each dimension m in making their classification judgments.¹ A geometric interpretation for the operation of the attention weights is that of stretching and shrinking the psychological space along its component dimensions (see Figure 2.1).

Finally, the similarity between item i and exemplar j is given by

$$s_{ij} = e^{-cd_{ij}^p}, \quad (3)$$

where c is a freely estimated sensitivity parameter that reflects the rate at which similarity declines with distance. When c is large, the similarity gradient is steep; that is, similarity falls off rapidly with increasing distance in the space. In this situation, the GCM acts very much like a nearest-neighbour classifier (i.e., one in which classification decisions are based primarily on the category membership of a test item's nearest neighbour). When c is small, the similarity gradient is shallow, and

¹ More precisely, the weights measure the span of each dimension *relative* to some prior, 'neutral' scaling or physical specification of the stimuli. To the extent that a prior scaling was obtained under conditions in which each dimension was equally relevant, then the weights estimated in a classification task can be interpreted in terms of the amount of 'attention' devoted to each dimension for purposes of the classification.

numerous exemplars may make major contributions to the classification decision. The value p in Equation 3 determines the shape of the function relating similarity to distance. In most cases, p is set equal to one, which yields an exponential relation between similarity and psychological distance (Shepard, 1987). In situations involving highly confusable stimuli, however, p is sometimes set equal to 2, yielding a Gaussian relation between similarity and distance (Nosofsky, 1985; for a theoretical interpretation, see Ennis, 1988).

In sum, the free parameters used for fitting the GCM to classification data are the sensitivity parameter c in Equation 3, the set of attention weights w_m in Equation 2; and the response-bias parameters and response-scaling parameter γ in Equation 1. (Because they can be constrained to sum to one, there are only $M - 1$ freely varying attention weights and $K_N - 1$ freely varying response-bias parameters.) All other quantities, including the x_{im} coordinate values (Equation 2), the V_{kK} memory strengths (Equation 1), the distance metric (value of r in Equation 2), and the similarity gradient (value of p in Equation 3) are fixed by the nature of the experimental design or are derived from independent sources.

Motivation

The GCM is a generalization of the original context model proposed by Medin and Schaffer (1978). Nosofsky (1984, 1986) integrated the Medin–Schaffer context model with classic theories developed in the areas of choice and similarity. This integration provided a firm theoretical foundation for the context model and allowed the original version to be extended to more widely varying classification paradigms. For example, the original context model was applied in highly simplified domains involving stimuli that varied along only binary-valued, separable dimensions. By comparison, most applications of the GCM are in continuous-dimension domains. In addition, the GCM is readily applied to predict classification performance involving integral-dimension stimuli as well as separable-dimension stimuli (Nosofsky, 1987). Moreover, when the GCM is used in combination with MDS approaches (in order to locate exemplars in psychological similarity spaces), it can be applied to predict classification performance in rich and complex domains in which the underlying stimulus dimensions may not be known (e.g., Shin & Nosofsky, 1992).

In the original version of the context model, Medin and Schaffer proposed what is essentially Equation 1 as a choice rule for classification. (Their equation did not include the memory-strength terms, bias

parameters, or the γ response-scaling parameter.) However, they did not provide a strong justification for the use of the choice rule beyond the fact that it seemed to work. Nosofsky (1984, 1986) provided a deeper foundation for the context-model response rule by noting a strong relation between it and the classic *similarity-choice model* (SCM) for predicting confusions in identification paradigms (Luce, 1963; Shepard, 1957). I briefly review this motivating relation here.

In an identification paradigm, there are n distinct stimuli, and each stimulus is assigned a unique response. The data are summarized in an $n \times n$ identification confusion matrix (see left panel of Figure 2.2), where each cell (i, j) of the matrix gives the conditional probability with which stimulus i is identified as stimulus j . The SCM is one of the leading descriptive models for predicting confusion probabilities in identification paradigms. According to the SCM, the probability with which stimulus i is identified as stimulus j is given by

$$P(j | i) = \frac{b_j s_{ij}}{\sum_{k=1}^n b_k s_{ik}}, \quad (4)$$

where b_j ($0 < b_j$) is the bias for response j , and where s_{ij} ($0 < s_{ij}$, $s_{ij} = s_{ji}$) is the similarity between stimuli i and j .

In a classification paradigm, the n stimuli are assigned to one of K_N categories ($K_N < n$). The data are summarized in an $n \times K_N$ confusion matrix, where cell (i, \mathcal{J}) of the matrix gives the conditional probability with which stimulus i is classified in Category \mathcal{J} . An illustration with $K_N = 2$ is provided in the right panel of Figure 2.2.

An intuitively compelling idea for predicting classification confusions from identification confusions was first proposed by Shepard, Hovland, and Jenkins (1961). The idea was that, in a classification paradigm, any time a stimulus is confused with another member of its own category, it would result in a correct classification response. Only between-class stimulus confusions would result in classification errors. Coined the *mapping hypothesis*² by Nosofsky (1986), the idea is illustrated schematically in Figure 2.2. In the figure, we imagine that stimuli 1–4 belong to Category A, whereas stimuli 5–8 belong to Category B. As illustrated in Figure 2.2, according to the mapping hypothesis, stimulus 3 would

² Identification paradigms involve a one-to-one mapping of stimuli onto responses. Classification paradigms involve a many-to-one mapping of stimuli onto responses. The mapping hypothesis allows one to use data from one-to-one mapping paradigms to predict performance in many-to-one mapping paradigms.

MAPPING HYPOTHESIS

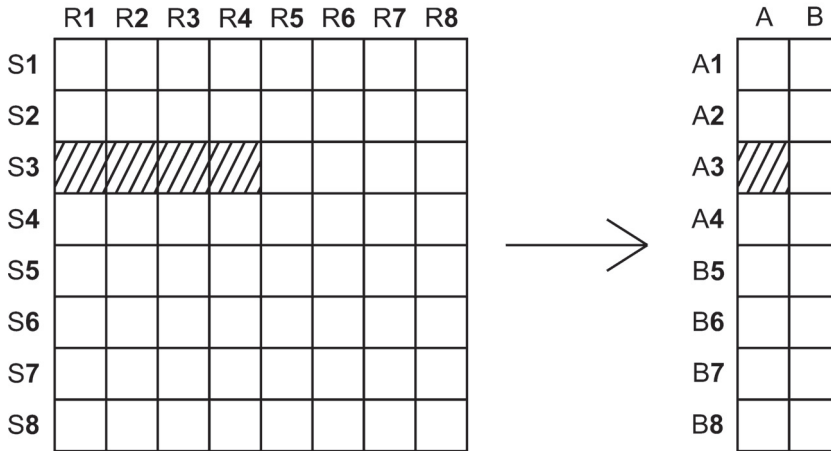


Figure 2.2 Left panel: an 8×8 stimulus-response (S-R) confusion matrix for an identification experiment. Right panel: an 8×2 S-R confusion matrix for a classification experiment (the same stimuli are used as in the identification task). Stimuli 1–4 are assigned to Category A, and stimuli 5–8 are assigned to Category B. According to the mapping hypothesis, one predicts the probability that stimulus 3 is classified in Category A by summing over the probabilities that stimulus 3 is identified as either stimulus 1, 2, 3, or 4 in the identification task. (From Nosofsky (1986), published by APA. Reprinted with permission.)

receive the correct response in the classification paradigm whenever it is confused with stimuli 1–4 in the identification paradigm.

More generally, if the mapping hypothesis is correct, then the probability that stimulus i is classified into Category \mathcal{J} would be found by summing over the probabilities that it is confused with any member of Category \mathcal{J} in the identification paradigm. It is straightforward to show that, if the SCM (Equation 4) describes these identification confusion probabilities, then the predicted classification probabilities are given by what is essentially Equation 1, i.e., the context-model response rule. The main difference is that because the response set has changed, the individual-item response bias parameters are replaced with category-level response-bias parameters.³

³ A further requirement for the mapping hypothesis to hold is that the γ response-scaling parameter be equal to one, which was true in the original formulation of the context model.

Although intuitively compelling, Shepard *et al.* (1961) observed systematic failures in using the mapping hypothesis to predict classification confusions on the basis of identification confusions. Although a full discussion goes beyond the scope of this chapter, Nosofsky (1984, 1986) proposed that the mapping hypothesis not be abandoned completely. The key idea was that the similarity parameters that operate in Equations 1 and 4 may change systematically across the identification and classification paradigms, because of the operation of the selective-attention processes depicted in Figure 2.1. Thus, Nosofsky (1986) proposed that rather than there being a *direct* mapping relation, a more abstract *indirect* mapping relation may connect identification and classification performance. According to this proposal, performance in both paradigms is governed by similarity relations among distinct exemplars, as formalized in Equations 1 and 4. However, these inter-exemplar similarity relations change systematically across the paradigms because of the operation of selective attention. Nosofsky (1984, 1986, 1987) provided strong support for this idea in theoretical and empirical work that investigated relations between identification and classification performance.

The motivation for the similarity equations used in the GCM (Equations 2 and 3) grows directly from decades of research and theoretical development in the area of similarity. MDS approaches to modelling similarity have long been among the most general and successful approaches in the field. Indeed, so successful have been those approaches that some of the discovered regularities have been proposed as candidates for universal laws of psychological generalization (Shepard, 1987). In particular, there is a great deal of support for the idea that psychological distance relations among integral-dimension stimuli are well described by embedding the stimuli in multidimensional Euclidean spaces, whereas psychological distance relations among separable-dimension stimuli are better approximated by a city-block distance metric. Furthermore, Shepard (1987) reviewed decades of research that point to an approximately exponential relation (Equation 3 with $p = 1$) between similarity, measured in terms of probability of stimulus generalization, and distance in these psychological spaces.

In their original formulation of the context model, Medin and Schaffer (1978) proposed an inter-dimensional multiplicative rule for computing stimulus similarity. The similarity between stimuli i and j was given by

$$s_{ij} = \prod_{m=1}^M s_m^{\delta_m(i,j)}, \quad (5)$$

where s_m ($0 \leq s_m \leq 1$) is a freely estimated dimension- m similarity parameter; δ is an indicator variable set equal to 1 if stimuli i and j mismatch on dimension m , and set equal to zero if they match on dimension m . Thus, the overall similarity between stimuli i and j is simply the product of their similarities along each of their mismatching dimensions. This inter-dimensional multiplicative rule is a special case of the MDS approach to modelling similarity that is embodied in Equations 2 and 3 of the GCM. In particular, an inter-dimensional multiplicative rule arises whenever $p = r$ in Equations 2 and 3 (see Nosofsky, 1986, p. 42, for further discussion). The particular highly constrained rule used by Medin and Schaffer (Equation 5) arises when, in addition, the stimuli vary along a set of independent, binary-valued dimensions. However, as explained above, use of the MDS approach in the GCM allows for a far more general application of the model to diverse classification domains.

Because of its combination of assumptions involving an exemplar-based category representation and a non-linear similarity rule, the GCM is highly sensitive to effects of specific exemplars and to within-class correlational structure in categories (for extensive discussion, see, for example, Medin and Schaffer, 1978; Nosofsky, 1992a). A variety of experiments have been reported for illustrating the importance of such sensitivity. For example, in some studies, category structures have been designed in which an individual stimulus i is more similar than is stimulus j to the prototype (central tendency) of their category; yet, stimulus j has high similarity to specific exemplars of the category, whereas stimulus i does not. In such studies, it is often found that subjects classify more accurately the stimulus with high exemplar-specific similarity than the stimulus with high similarity to the prototype (for reviews and examples, see Nosofsky, 1992a, 2000; Nosofsky & Zaki, 2002). Exemplar models with non-linear similarity rules (such as the GCM) account naturally for such effects, whereas major alternative models, such as prototype models, do not.

Implementation recommendations

In the present section I present some practical advice on implementing the GCM. The key free parameters that are almost always estimated are the sensitivity parameter (c) in Equation 3 and the attention-weight parameters (the w_m) in Equation 2. An interesting working hypothesis is that, with learning, the participant may adopt a set of ideal-observer weights, for example, a set of weights that would allow the participant to maximize his or her percentage of correct classifications. Thus, it is interesting to compare the best-fitting attention-weight parameters to

the ideal-observer weights. In many situations, the best-fitting weights turn out to approximate the ideal-observer weights (e.g., Nosofsky, 1984, 1986, 1991b).

Unless there are strong experimental manipulations involving differential category payoffs or frequencies, the response-bias parameters tend not to contribute substantially to the model fits, and can generally be set equal to one. The γ response-scaling parameter (Equation 1) is important when modelling performance at the individual-observer level, to allow the model to capture the deterministic response strategies that individual observers sometimes use (e.g., McKinley & Nosofsky, 1995; Nosofsky & Zaki, 2002). Practical experience suggests, however, that in fitting averaged data, γ can be held fixed at one without much loss in predictive accuracy.

As discussed in the introduction, in situations involving fairly discriminable stimuli, an exponential decay function for relating similarity to distance ($p = 1$ in Equation 3) is always assumed. For integral-dimension stimuli, a Euclidean metric ($r = 2$) is assumed in Equation 2; whereas for highly separable-dimension stimuli, a city-block metric ($r = 1$) is assumed. By contrast, in situations involving highly confusable perceptual stimuli, the values $p = 2$ and $r = 2$ tend to provide much better fits. (These parameter settings are probably reflecting extensive perceptual noise in the stimulus representations that exists in high-confusability situations – see Ennis (1988) for further discussion.)

The memory-strength values (Equation 1) are generally not treated as free parameters. Instead, they are usually set proportional to the relative frequency with which each stimulus is presented in combination with associated category feedback during the training phase of the experiment. In cases in which one models trial-by-trial performance data, a memory-strength value is often attached to each individual exemplar presented on each trial; in this case, the memory strengths are assumed to decay exponentially with lag of presentation of the exemplars (e.g., McKinley & Nosofsky, 1995).

In general, the GCM is fitted to classification choice-probability data by using a maximum-likelihood criterion, under the assumption that the distribution of responses into categories is multinomial in form (see Nosofsky, 1989, for examples). Hierarchically nested versions of the model, in which some parameters are held fixed at default values, can be tested against the full version of the model by using standard likelihood-ratio techniques (see Nosofsky, 1989, for examples) or alternative methods such as AIC or BIC that penalize models for their number of free parameters. Because analytic solutions for the maximum-likelihood parameters are generally not available except in exceedingly simple cases,

computer search is used for locating the best-fitting parameters. As is the case for fitting any highly nonlinear model to data, multiple starting configurations should be used in the parameter searches to guard against local minima.

The GCM has been a highly successful and widely applied model. In addition, its free parameters are easily interpretable and provide measurements of psychological processes of fundamental interest. For example, researchers may be interested in the extent to which different populations of subjects adopt alternative patterns of selective attention to the dimensions that compose a set of stimuli (e.g., Viken *et al.*, 2002). The derived attention-weight parameters from the model provide such information. For these reasons, an important recent development is the availability of a general-purpose computer package for conducting Bayesian analyses of the GCM (Vanpaemel, 2009; see also Lee, 2008). The package provides estimates of the posterior distributions of the model's parameters to allow for easy inference and interpretation of the effects of different experimental manipulations on psychological processing.

Application to an example

An example application of the GCM to a previously published data set (Nosofsky, 1989) is briefly reviewed in Figure 2.3 and Table 2.1. Because the fits are to averaged data, they should be considered as merely illustrative. The stimuli were a set of semicircles with an embedded radial line. The semicircles varied orthogonally in their size (four levels) and in the angle or orientation of the radial line (four levels) to create a 16-member stimulus set. Subjects were tested in four different categorization conditions, which are illustrated schematically in Figure 2.3. In the figure, the columns of each 4×4 grid correspond to the different levels of angle, and the rows correspond to the different levels of size. Cells that are marked in their centre with a boldface 1 represent training exemplars of Category 1, whereas cells marked with a 2 were training exemplars of Category 2. Unmarked cells were unassigned transfer stimuli. During an initial training phase, subjects were presented with only the training exemplars and were provided with trial-by-trial corrective feedback. During a subsequent test phase, subjects were presented with both training and transfer stimuli. Feedback continued to be provided on trials in which training exemplars were presented, but was withheld on trials in which unassigned transfer stimuli were presented.

As can be seen in the figure, to provide evidence of generality, a variety of different category structures were tested. The Size and Angle categorizations provide examples of 'unidimensional' category structures, in

Table 2.1 *Maximum-likelihood parameters and summary fits for full and restricted versions of the exemplar-similarity categorization model (reprinted from Nosofsky, 1989)*

Condition	Model	Parameters			Fits		
		c	w_1	b_1	SSE	%Var	$-\ln L$
Size	Unconstrained	1.60	0.10	0.50	0.015	99.4	40.8
	Equal attention	2.38	(0.50)	0.49	0.077	97.0	72.0
	Equal bias	160	0.10	(0.50)	0.015	99.4	40.8
Angle	Unconstrained	3.20	0.98	0.43	0.010	99.6	44.3
	Equal attention	3.57	(0.50)	0.45	0.305	86.4	164.3
	Equal bias	3.09	1.00	(0.50)	0.029	98.7	56.8
Criss-cross	Unconstrained	1.62	0.80	0.45	0.025	95.2	47.7
	Equal attention	1.23	(0.50)	0.45	0.087	83.1	64.6
	Equal bias	3.00	0.93	(0.50)	0.046	91.1	56.7
Diagonal	Unconstrained	2.42	0.81	0.49	0.023	98.4	483
	Equal attention	1.81	(0.50)	0.48	0.217	85.0	109.4
	Equal bias	2.42	81	(0.50)	0.021	98.6	49.1

Values in parentheses were constrained a priori. The parameter w_1 gives the attention weight for Angle, and $1 - w_1$ the attention weight for Size. SSE is the sum of squared deviations between predicted and observed Category 1 probabilities; %Var is the percentage of variance accounted for; $\ln L$ is the log likelihood.

which one dimension is far more relevant than is the other for purposes of classification. These structures were tested to provide clear evidence of the role of selective attention in classification. Consider, for example, the Angle categorization. Note that stimulus 14 is an unassigned transfer stimulus. In terms of ‘overall similarity’, stimulus 14 is more similar to the exemplars of Category 2 than to the exemplars of Category 1 (it lies extremely close to exemplar 15 of Category 2). However, if subjects attend selectively to the relevant angle dimension, then the space will be stretched along the horizontal angle dimension, and shrunk along the vertical size dimension, rendering stimulus 14 more similar to the exemplars of Category 1. An analogous situation holds for stimulus 9 of the Size categorization. The Criss-Cross categorization provides an example of a continuous-dimension biconditional structure. Note that the structure is non-linearly separable, i.e., the members of the contrasting categories cannot be separated by drawing a straight line through the space. Whereas various categorization models are applicable only in cases involving linearly separable categories, the GCM applies generally regardless of the form of the category structure. Finally, the Diagonal

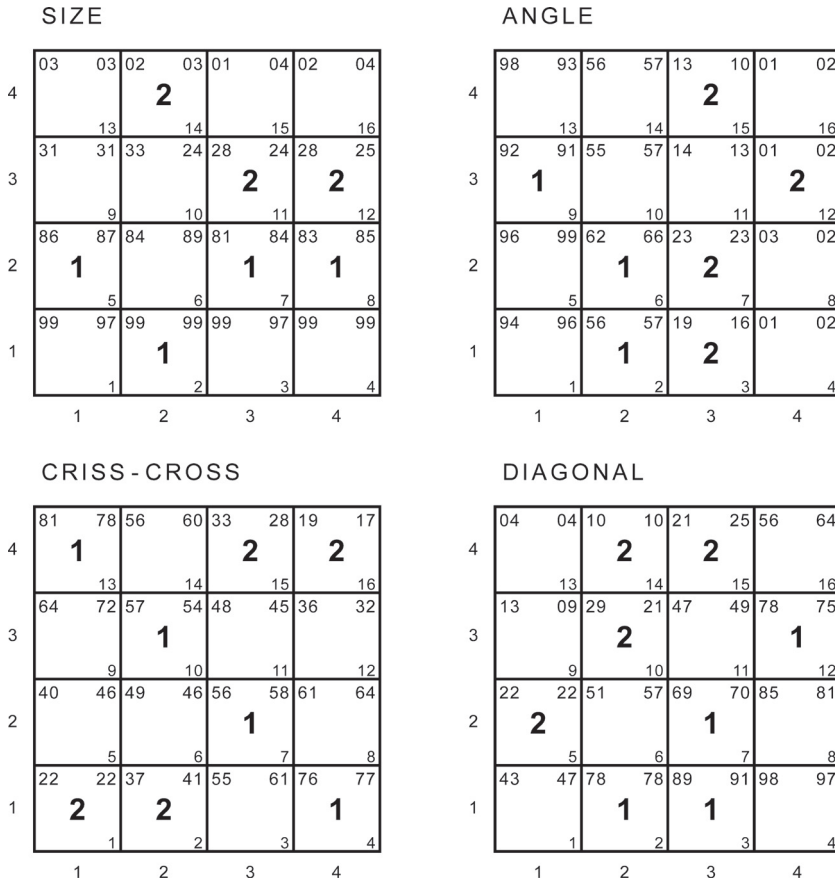


Figure 2.3 Schematic illustration of the four category structures tested by Nosofsky (1989). Cells marked with a 1 (2) depict training exemplars of Category 1 (2). Unmarked cells depict unassigned transfer stimuli. Top-left value in each cell is the predicted Category-1 response probability from the GCM. Top-right value in each cell is the observed Category-1 response probability. (Reprinted from Nosofsky, 1989.)

categorization provides an example of a fairly ‘natural’ two-dimensional category structure in which the exemplars of contrasting categories can be separated by drawing a diagonal line through the space.

In a preliminary similarity-scaling study involving the collection of identification confusion data (see Nosofsky, 1989, for details), a

two-dimensional scaling solution was derived for the set of 16 stimuli. Not surprisingly, the derived two-dimensional solution closely reflected the 4×4 grid structure of the stimulus set (see Nosofsky, 1989, Figure 2.3). This two-dimensional scaling solution was then used in combination with the GCM to predict the classification data obtained in each of the four categorization conditions. The free parameters in the ‘full’ version of the model were the sensitivity parameter c , the dimension-1 attention weight w_1 (with $w_2 = 1 - w_1$), and the Category-1 response-bias parameter b_1 (with $b_2 = 1 - b_1$). All other parameters were set at the default values described previously in this chapter. (Because the stimuli were highly confusable, the Gaussian-Euclidean version of the model was used, with $p = 2$ and $r = 2$.) Parameters were estimated for each individual categorization condition that provided a maximum-likelihood fit to the data. In addition, special cases of the model were also fit to the data. In one special case, the attention weights were held fixed at $w_1 = w_2 = 0.5$; and in a second special case, the bias parameters were held fixed at $b_1 = b_2 = 0.5$. By comparing the fits of the special-case versions to the full version of the GCM, one gains evidence regarding the importance of the various free parameters.

The results of the model fitting are reported in Figures 2.3 and 2.4 and in Table 2.1. In Figure 2.3, the top-right value in each cell gives the observed probability with which subjects classified the stimulus into Category 1, whereas the top-left value gives the predicted probability from the GCM. A summary of all of these observed-against-predicted probabilities for all four categorization conditions is provided in the scatterplot in Figure 2.4. It can be seen from inspection that the model provides extremely accurate predictions of the data in all four conditions. The quantitative summary fits from the model are also reported in Table 2.1. Although the criterion of fit was to maximize likelihood (or, equivalently, to minimize the negative log-likelihood), the table also reports the sum-of-squared deviations between the predicted and observed classification probabilities, as well as the percentage of variance accounted for in each condition.

Finally, the table reports the fit results for the special-case models described above. Of greatest interest are the results for the special-case version that assumes equal attention to the two stimulus dimensions. In all conditions, this special-case model fits significantly worse than does the full version (see Nosofsky, 1989, for detailed reports of likelihood-ratio tests that compare the full version to the special cases). These results provide clear evidence of the need to incorporate assumptions about selective attention into the modelling. They can be understood most easily by considering the results for the critical transfer stimuli

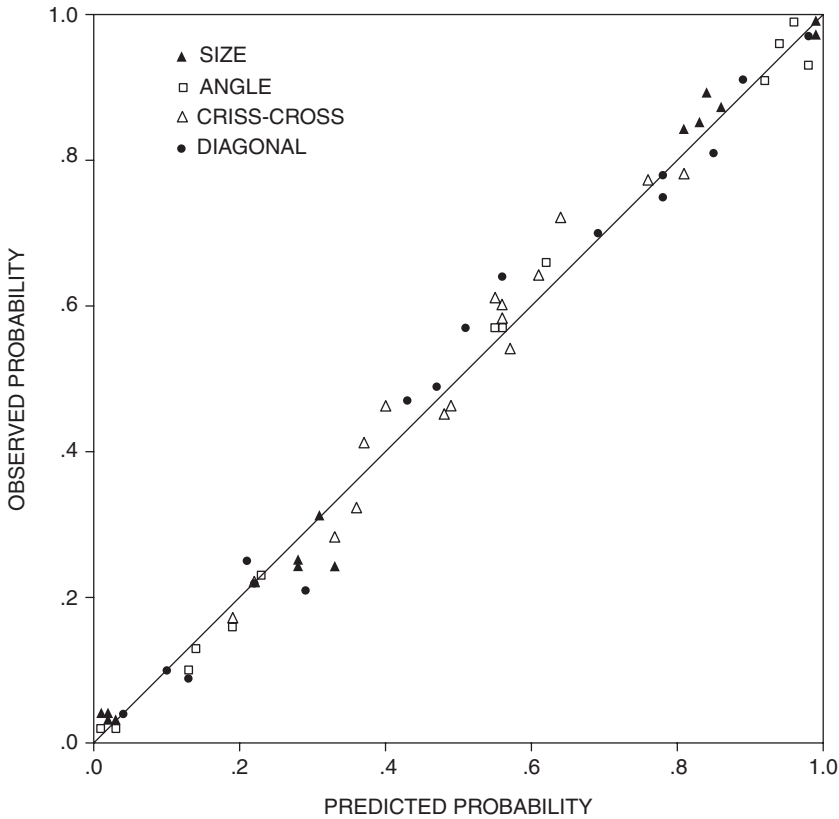


Figure 2.4 Scatterplot of observed against GCM-predicted Category-1 response probabilities for each stimulus in all four categorization conditions. (Reprinted from Nosofsky, 1989.)

described above. For example, in the Angle categorization, the equal-attention model predicted that transfer stimulus 14 would be classified into Category 2 with high probability, because it has greater overall similarity to the exemplars of Category 2 than to Category 1. However, the full version of the model, which allows for selective attention to the relevant angle dimension, predicts correctly that transfer stimulus 14 is classified with somewhat higher probability into Category 1. Indeed, the maximum-likelihood estimate for the attention-weight parameter in the Angle categorization is $w_1 = 0.98$ (see Table 2.1), providing clear evidence for the operation of selective attention in this condition.

Whereas the role of the selective-attention parameters in the GCM has been emphasized and studied systematically in past research, much less

emphasis has been placed on the role of the response-bias parameters. However, in the same way that certain patterns of selective attention may yield better classification performance than others (depending on the category structure), so may certain patterns of response bias yield better performance than others (for examples in situations involving categories with differing variabilities, see Cohen, Nosofsky, & Zaki, 2001). Thus, to the extent that human observers are adaptive and adjust parameter settings to achieve task goals, it would not be surprising to find evidence of systematic shifts in patterns of category response bias as a function of experimental conditions.

Relations and extensions

Although a detailed discussion goes beyond the scope of this chapter, it is important to remark that the GCM is closely related to a variety of rational, Bayesian, and powerful statistical/machine-learning models of categorization (e.g., Anderson, 1990; Ashby & Alfonso-Reese, 1995; Jakel, Scholkopf, & Wichman, 2008; Nosofsky, 1990, 1991a; Shepard, 1987). Thus, it is interesting to speculate that exemplar-based classification schemes evolved because they provide highly adaptive and flexible solutions to the goals of classification given the structure of natural categories and limits on the computational complexity of the mind (Anderson, 1990). Furthermore, perhaps because the GCM builds upon classic past models in the fields of similarity and classification, and it has itself been a highly successful and widely applied model, there have also been numerous extensions of the GCM. In this final section I briefly review some of these extensions.

One extension is Kruschke's (1992) highly influential ALCOVE (attention-learning covering-map) model. In essence, ALCOVE adopts the GCM's exemplar-based representational assumptions, its MDS-based similarity assumptions, and its assumptions about selective attention, and embeds them within the framework of a connectionist learning network. A key potential advantage is that, whereas the selective-attention weights are free parameters in the GCM, ALCOVE provides a mechanism that *learns* the attention weights on a trial-by-trial basis. In addition, in standard applications of the GCM, the memory strengths associated with individual exemplars are assumed simply to be proportional to the frequency with which each exemplar is presented in combination with given category feedback. By contrast, in ALCOVE, learning is error driven, and association weights develop between exemplars and categories that are more intricate and potentially more adaptive than those allowed by the GCM.

Whereas the GCM is limited to predicting choice-probability data in classification, other major extensions of the model enable it also to predict classification response times (RTs). Nosofsky and Palmeri's (1997) *exemplar-based random-walk* (EBRW) model adopts the same representational assumptions as does the GCM. However, classification decision-making is governed by a random-walk process. The random walk is driven by exemplars that are retrieved from memory. Importantly, in a two-category experiment, the GCM response-rule (Equation 1) emerges as a special case of the EBRW model, so the EBRW provides a process-level interpretation for the GCM. It also predicts successfully a wide variety of fundamental effects involving classification RT data, including effects of distance-from-boundary, familiarity, practice, and probabilistic feedback (Nosofsky & Palmeri, 1997; Nosofsky & Stanton, 2005). Another major extension is Lamberts' (2000) *extended generalized context model for response times* (EGCM-RT). This extension also adopts the GCM's exemplar-based representational and similarity assumptions. However, it assumes that categorization involves the gradual construction of a perceptual representation through a process of information accumulation. In particular, the perceptual representation of a stimulus is gradually built through a process of stochastic, parallel sampling of the stimulus's individual dimensions. Classification RTs are determined by the length of time with which the sampling process operates, which in turn is closely related to outputs from the GCM response rule.

A long-standing debate in the classification literature has involved a contrast between exemplar and prototype models. Whereas exemplar models assume that categories are represented in terms of individually stored exemplars, prototype models assume instead a single summary representation that is the central tendency of the individual exemplars. In Chapter 3 of this volume, Minda and Smith consider prototype models. They argue that prototype models provide better accounts of data than do exemplar models in a variety of experiments involving visual categories. However, this conclusion has been disputed in numerous studies (see e.g., Busemeyer, Dewey, & Medin, 1984; Nosofsky, 2000; Nosofsky & Zaki, 2002; Palmeri & Flanery, 2002; Rehder & Hoffman, 2005; Stanton, Nosofsky, & Zaki, 2002; Zaki & Nosofsky, 2004, 2007; Zaki *et al.*, 2003). An intermediate view is that multiple prototypes may be formed to represent a category, where a varying number of exemplars may be averaged together to form each individual prototype. In the *varying abstraction model* (VAM) of Vanpaemel and Storms (2008), all possible multiple-prototype representations are considered, and the version that fits best is taken as the underlying category representation. Although the VAM is far more complex than the GCM, results from

Vanpaemel and Storms suggest that the added complexity is justified in terms one's ability to accurately predict human classification performance. A closely related alternative idea is the Rex Leopold I model of De Schryver, Vandist, and Rosseel (2009). This model posits that categories are not represented in terms of the complete set of experienced exemplars, but in terms of reduced subsets of the exemplars. Analogous to the VAM, all possible reduced subsets are considered, and the best-fitting subset is taken as the category representation.

Another important extension of the GCM is Stewart and Brown's (2005) *similarity-dissimilarity exemplar model*. This model posits that the evidence in favour of a category is not based solely on the similarity of a test item to the exemplars of that category, but on its *dissimilarity* to the exemplars of contrast categories. This extended model accounts successfully for the use of difference information and category-contrast effects in classification. A final example is Pothos and Bailey's (2009) extension of the GCM to judgments of category intuitiveness and unsupervised categorization. This unsupervised GCM operates by computing, for any given partitioning of exemplars into categories, the overall prediction error associated with that partitioning. The category assignment that results in the smallest prediction error is considered to be the most intuitive.

Finally, although I have focused on the issue of categorization in this chapter, perhaps the most important achievement of the GCM is that essentially the same model has been used to account for varieties of other fundamental cognitive processes, including individual-item identification, the development of automaticity, and old-new recognition performance (e.g., Nosofsky, 1986, 1987, 1988, 1991b; Nosofsky & Zaki, 1998; Nosofsky & Stanton, 2006; Palmeri, 1997). The successful applications of the GCM in these domains, and the use of the model to explain relations between categorization and performance in other fundamental tasks, suggests the possibility of developing unified, exemplar-based theories of cognitive representation and processing.

REFERENCES

- Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: LEA.
- Ashby, F. G., & Alfonso-Reese, L. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216–233.
- Ashby, F. G., & Maddox, W. T. (1993). Relations between exemplar, prototype, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.
- 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, and Cognition*, **10**, 638–648.
- Caroll, J.D., & Wish, M. (1974). Models and methods for three-way multi-dimensional scaling. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (eds.), *Contemporary Developments in Mathematical Psychology* (Vol. 2). San Francisco: W.H. Freeman.
- Cohen, A.L., Nosofsky, R.M., & Zaki, S.R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, **29**, 1165–1175.
- De Schryver, M., Vandist, K., & Rosseel, Y. (2009). How many exemplars are used? Explorations with the Rex Leopold I model. *Psychonomic Bulletin & Review*, **16**, 337–343.
- Ennis, D.M. (1988). Confusable and discriminable stimuli: comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, **117**, 408–411.
- Garner, W.R. (1974). *The Processing of Information and Structure*. New York: Wiley.
- Jakel, F., Scholkopf, B., & Wichman, F.A. (2008). Generalization and similarity in exemplar models of categorization: insights from machine learning. *Psychonomic Bulletin & Review*, **15**, 256–271.
- Kruschke, J.K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22–44.
- Lamberts, K. (2000). Information accumulation theory of categorization response times. *Psychological Review*, **107**, 227–260.
- Lee, M.D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, **15**, 1–15.
- Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (eds.), *Handbook of Mathematical Psychology* (pp. 103–189). New York: Wiley.
- 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 and Performance*, **21**, 128–148.
- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207–238.
- Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 104–114.
- (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, **38**, 415–432.
- (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39–57.
- (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 87–109.
- (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 700–708.

- (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics*, **45**, 279–290.
- (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, **34**, 393–418.
- (1991a). Relation between the rational model and the context model of categorization. *Psychological Science*, **2**, 416–421.
- (1991b). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 3–27.
- (1992a). Exemplars, prototypes, and similarity rules. In A. F. Healy & S. M. Kosslyn (eds.), *Essays in Honor of William K. Estes, Vol. 1: From Learning Theory to Connectionist Theory* (pp. 149–167). Hillsdale, NJ: LEA.
- (1992b). Similarity scaling and cognitive process models. *Annual Review of Psychology*, **43**, 22–53.
- (2000). Exemplar representation without generalization: Comment on Smith and Minda's (2000) "Thirty categorization results in search of a model". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**, 1735–1743.
- Nosofsky, R.M., & Palmeri, T.J. (1997). An exemplar-based random-walk model of speeded classification, *Psychological Review*, **104**, 266–300.
- Nosofsky, R.M., & Stanton, R.D. (2005). Speeded classification in a probabilistic category structure: contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance*, **31**, 608–629.
- (2006). Speeded old-new recognition of multidimensional perceptual stimuli: modeling performance at the individual-participant and individual-item levels. *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 314–334.
- Nosofsky, R.M., & Zaki, S.R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychological Science*, **9**, 247–255.
- (2002). Exemplar and prototype models revisited: response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **28**, 924–940.
- Palmeri, T.J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23**, 324–354.
- Palmeri, T.J., & Flanery, M.A. (2002). Memory systems and perceptual categorization. In B.H. Ross (ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 141–189). San Diego, CA: Academic Press.
- Pothos, E.M., & Bailey, T.M. (2009). Predicting category intuitiveness with the rational model, the simplicity model, and the Generalized Context Model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**, 1062–1080.
- Rehder, B., & Hoffman, A.B. (2005). Thirty-something categorization results explained: selective attention, eyetracking, and models of category learning.

- Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**, 811–829.
- Shepard, R.N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325–345.
- (1958). Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, **55**, 509–523.
- (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54–87.
- (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317–1323.
- Shepard, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75** (13), Whole No. 517.
- Shin, H.J., & Nosofsky, R.M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, **121**, 278–304.
- Stanton, R.D., Nosofsky, R.M., & Zaki, S. R. (2002). Comparisons between exemplar similarity and mixed prototype models using a linearly separable category structure. *Memory & Cognition*, **30**, 934–944.
- Stewart, N., & Brown, G.D.A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, **49**, 403–409.
- Vanpaemel, W. (2009). BayesGCM: software for Bayesian inference with the Generalized Context Model. *Behavior Research Methods*, **41**, 1111–1120.
- Vanpaemel, W., & Storms, G. (2008). In search of abstraction: the varying abstraction model of categorization. *Psychonomic Bulletin & Review*, **15**, 732–749.
- Viken, R.J., Treat, T.A., Nosofsky, R.M., McFall, R.M., & Palmeri, T.J. (2002). Modeling individual differences in perceptual and attentional processes related to bulimic symptoms. *Journal of Abnormal Psychology*, **111**, 598–609.
- Zaki, S. R., & Nosofsky, R. M. (2004). False prototype enhancement effects in dot pattern categorization. *Memory & Cognition*, **32**, 390–398.
- (2007). A high-distortion enhancement effect in the prototype learning paradigm: dramatic effects of category learning during test. *Memory & Cognition*, **35**, 2088–2096.
- Zaki, S.R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: a reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **29**, 1160–1173.