Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 6, 1160–1173

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.6.1160

Prototype and Exemplar Accounts of Category Learning and Attentional Allocation: A Reassessment

Safa R. Zaki
Williams College

Robert M. Nosofsky, Roger D. Stanton, and Andrew L. Cohen
Indiana University Bloomington

In a recent article, J. P. Minda and J. D. Smith (2002) argued that an exemplar model provided worse quantitative fits than an alternative prototype model to individual subject data from the classic D. L. Medin and M. M. Schaffer (1978) 5/4 categorization paradigm. In addition, they argued that the exemplar model achieved its fits by making untenable assumptions regarding how observers distribute their attention. In this article, we demonstrate that when the models are equated in terms of their response-rule flexibility, the exemplar model provides a substantially better account of the categorization data than does a prototype or mixed model. In addition, we point to shortcomings in the attention-allocation analyses conducted by J. P. Minda and J. D. Smith (2002). When these shortcomings are corrected, we find no evidence that challenges the attention-allocation assumptions of the exemplar model.

A classic issue in the categorization literature has been whether people represent categories in terms of abstracted prototypes or in terms of specific exemplars. According to prototype models, people represent categories in terms of some central tendency computed over the category training instances and classify objects on the basis of how similar they are to the prototypes of the alternative categories (Homa & Vosburgh, 1976; Posner & Keele, 1968; Reed, 1972). By contrast, according to exemplar models, people represent categories by storing the individual training instances themselves (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986).

A well-known experimental paradigm that has been used for contrasting the predictions of exemplar and prototype models is the Medin and Schaffer (1978) *5/4 category structure*, which is listed in Table 1.[1] In this paradigm, the stimuli are simple perceptual forms that vary along four salient binary-valued dimensions. The stimuli are divided into two categories. The logical values of the prototype of Category A are assumed to be 0 0 0 0, and the logical values of the prototype of Category B are assumed to be 1 1 1 1. Subjects are trained on the first nine items and are then given a transfer test that includes all the items in the list. This category structure is diagnostic because prototype and exemplar models tend to make opposite predictions for specific items. Most critically, prototype models predict that people will perform better on Stimulus A1 than on Stimulus A2 because A1 shares more features with its category prototype. In contrast, exemplar models generally predict an A2 advantage because A2 is highly similar to (i.e., shares three features with) two Category A exemplars and no Category B exemplars. In fact, the A2 advantage has been observed in numerous studies. Furthermore, when exemplar and prototype models are fitted to the classification data in this design, the results generally favor the predictions from the exemplar model (for reviews, see Nosofsky, 1992, 2000; but see Smith & Minda, 2000, for an opposing viewpoint).

However, whereas previous research concentrated on fits of the models to aggregate data, Minda and Smith (2002) recently collected data from the 5/4 category structure but fit the models to the individual subject data. Their research revealed two potentially problematic issues for the exemplar model. First, in contrast to previous results, the exemplar model provided worse fits to the data than the prototype model. This result was obtained even though the A2 advantage was consistently observed in Minda and Smith's (2002) data sets. A second problematic issue involved certain attention-allocation analyses conducted by Minda and Smith (2002). These researchers argued that, even to the extent that the exemplar model could achieve good fits, it did so by adjusting its attention-weight parameter estimates in untenable ways.

In this article we argue that there are several reasons for concern regarding the analyses and conclusions of Minda and Smith (2002). First, as we discuss in detail in the following sections, Minda and Smith (2002) tested only a restricted version of an

Safa R. Zaki, Department of Psychology, Williams College; Robert M. Nosofsky, Roger D. Stanton, and Andrew L. Cohen, Department of Psychology, Indiana University Bloomington.

We thank J. Paul Minda for providing us with the data from Experiment 2 of Minda and Smith (2002). This work was supported by National Institute of Mental Health Grant R01 MH48494.

Correspondence concerning this article should be addressed to Safa R. Zaki, Department of Psychology, Bronfman Science Center, Williams College, Williamstown, Massachusetts 01267. E-mail: szaki@williams.edu

[1] Our article addresses a replication of the Medin and Schaffer (1978) paradigm reported by Minda and Smith (2002). In their article, Minda and Smith reordered the dimension values of the Medin and Schaffer (1978) stimuli, such that Dimensions 1, 2, 3, and 4 in the Minda and Smith (2002) article refer to Dimensions 4, 3, 2, 1 in the Medin and Schaffer paradigm. For ease of comparison to Minda and Smith (2002), we use the reordered dimension labels.

Table 1
Medin and Schaffer (1978) 5/4 Category Structure

Stimulus	Dimension 1	2	3	4
Category A				
A1	1	0	0	0
A2	1	0	1	0
A3	0	1	0	0
A4	0	0	1	0
A5	0	0	0	1
Category B				
B1	1	1	0	0
B2	1	0	0	1
B3	0	1	1	1
B4	1	1	1	1
Transfer (T)				
T1	0	1	1	0
T2	1	1	1	0
T3	0	0	0	0
T4	1	1	0	1
T5	0	1	0	1
T6	0	0	1	1
T7	1	0	1	1

exemplar model. In particular, they tested a version of the generalized context model (GCM; Nosofsky, 1984, 1986) that assumes a probability-matching response rule. However, there have been numerous previous demonstrations that the GCM cannot fit individual subject data under this assumption because individual subjects tend to be more deterministic in their response strategies (Ashby & Gott, 1988; Ashby & Maddox, 1993; McKinley & Nosofsky, 1995; Nosofsky & Zaki, 2002). In contrast, the version of the prototype model that Minda and Smith (2002) tested could accommodate any level of probabilistic versus deterministic responding that is desired (Nosofsky & Zaki, 2002). In this article we demonstrate that when a full version of the GCM is applied, which is not constrained to predict probability matching, it achieves fits to Minda and Smith's (2002) data that are substantially better than those of the prototype model. Second, we argue in this article that there are a number of problems with the attention analyses that Minda and Smith (2002) conducted. When these problems are corrected, we find no evidence that challenges the attention-allocation assumptions made by the exemplar model.

Review of the Models

In this section we provide a brief review of the formal models that guide the present research.

The Generalized Context Model

According to the GCM (Nosofsky, 1984, 1986), people represent categories by storing individual exemplars in memory. The exemplars are represented as points in a multidimensional space. The stimuli in the Medin and Schaffer (1978) 5/4 structure varied along four binary-valued separable dimensions. Therefore, the psychological distance between two stimuli is calculated by using the weighted city-block metric,

$$d_{ij} = \sum_{m=1}^{4} w_m |x_{im} - x_{jm}|, \qquad (1)$$

where x_{im} is the value (either 0 or 1) of item *i* on dimension *m*, and $w_m (0 \leq w_m \leq 1, \sum w_m = 1)$ is the attention weight given to dimension *m*. The similarity between two stimuli is an exponential decay function of their distance,

$$s_{ij} = e^{-cd_{ij}}, \qquad (2)$$

where *c* is an overall sensitivity parameter (Shepard, 1987). Finally, the probability that item *i* is classified into Category A is given by

$$P(A|i) = \frac{[\sum s_{ia}]^{\gamma}}{[\sum s_{ia}]^{\gamma} + [\sum s_{ib}]^{\gamma}}, \qquad (3)$$

where $\sum s_{ia}$ and $\sum s_{ib}$ are the summed similarities of item *i* to the exemplars of Categories A and B, respectively, and γ is a response-scaling parameter (Ashby & Maddox, 1993; McKinley & Nosofsky, 1995). When $\gamma = 1$, the model assumes that subjects probability match to the relative summed similarities. At higher levels of γ, the model assumes that subjects respond more deterministically with the category that yields the larger summed similarity. This full version of the GCM has five free parameters: the sensitivity parameter *c*, the response-scaling parameter γ, and three freely varying attention weights (w_m).

Multiplicative Prototype Model

Like Minda and Smith (2002), we also tested a multiplicative prototype model (MPM), that is, a version of the prototype model that uses the same similarity functions as the GCM (Estes, 1986; Nosofsky, 1987, 1992). In this model, the distance between item *i* and Prototype A is given by

$$d_{iA} = \sum_{m=1}^{4} w_m \cdot |x_{im} - P_{Am}|, \qquad (4)$$

where P_{Am} denotes the value of Prototype A on dimension *m*, and the w_m are the attention weights. The similarity between item *i* and Prototype A is then given by

$$s_{iA} = e^{-cd_{iA}}, \qquad (5)$$

where *c* is the sensitivity parameter. The probability with which item *i* is classified into Category A is given by

$$P(A|i) = \frac{s_{iA}^{\gamma}}{s_{iA}^{\gamma} + s_{iB}^{\gamma}}. \qquad (6)$$

However, in the MPM, the γ response-scaling parameter cannot be estimated separately from the sensitivity parameter *c* (see Nosofsky & Zaki, 2002, for extensive discussion). Therefore, without loss of generality in this model, the γ parameter is set at 1. The MPM has four free parameters: the sensitivity parameter *c* and three freely varying attention weights (w_m).

Mixed-Prototype Model

In their mixed-prototype model of categorization, Smith and Minda (1998, 2000; Minda & Smith, 2001; Smith, Murray, & Minda, 1997) hypothesized that subjects store memories of particular exemplars after extended training. The stored exemplars, however, are assumed to be used only for the purpose of classifying the training instances themselves, with generalization to new items based solely on similarity to the prototypes. According to this mixed model, the probability that an old item from Category A is classified into Category A is given by

$$P(A|i) = e + (1 - e)\cdot\frac{g}{2} + (1 - e)\cdot(1 - g)\cdot proto(A|i), \qquad (7)$$

where e $(0 \leq e \leq 1)$ is the probability that the observer uses item-specific exemplar memories, g $(0 \leq g \leq 1)$ is a guessing probability, and $proto(A|i)$ is the probability with which item i is classified into Category A by a prototype process. If the item i was not presented during training, the parameter e is set to 0, and the model reduces to a categorization decision based solely on guessing and prototype-based similarity. This mixed model has six free parameters: sensitivity c, guessing g, exemplar-memory e, and three freely varying attention weights (w_m).

Modeling the Minda and Smith (2002) Datasets

Background

Minda and Smith (2002) fitted a version of the GCM as well as the MPM to the individual-subject data that they collected in the Medin and Schaffer (1978) 5/4 paradigm. Although the models could not be distinguished quantitatively on the basis of their fits to the transfer test data, the MPM fit the training data better than the GCM in two separate experiments. Minda and Smith (2002) interpreted these findings as an indication that the observers were using a prototype process in this task.

However, in modeling these data, Minda and Smith (2002) fitted only a version of the exemplar model in which the response-scaling parameter (γ) was set to 1. This is the value that was assumed in the original version of the context model (Medin & Schaffer, 1978) and in the early articles that first introduced the GCM (Nosofsky, 1984, 1986, 1987). Since then, however, it has been widely acknowledged that when this restricted model is applied to data at the individual subject level, it fails because subjects respond more deterministically than is predicted by a probability-matching rule (McKinley & Nosofsky, 1995; Nosofsky, 1991; Nosofsky & Johansen, 2000; Nosofsky & Palmeri, 1997; Nosofsky & Zaki, 2002). Therefore, a major aim of the present research was to test the full GCM (with γ allowed to vary) on its ability to account for Minda and Smith's (2002) individual-subject data.

In addition, in numerous previous articles, Smith and colleagues (Minda & Smith, 2001; Smith & Minda, 1998, 2000; Smith et al., 1997) have argued for the importance of the mixed-prototype model. They have suggested that following extended training on a fixed set of exemplars, subjects use the exemplar-memorization process described in our Review of the Models section. However, Minda and Smith (2002) did not consider the predictions of the mixed-prototype model in their recent replication of the Medin and Schaffer (1978) paradigm. Therefore, a second aim of the present research was to test the mixed-prototype model as well.

Data Set

The data that we used to assess the models were from Experiment 2 of Minda and Smith (2002). (Experiment 2 replicated their Experiment 1 but also included a transfer phase following initial training.) To briefly summarize, in their experiment, Minda and Smith (2002) trained subjects on the first nine stimuli listed in Table 1. Subjects were trained for 40 blocks of nine trials each. Each training stimulus appeared once in a random order in each block. Following training, subjects were shown both old and new items (Stimuli 1–16) in a transfer phase. There was a total of 8 blocks of 16 transfer trials, with each transfer stimulus appearing once in a random order in each block.

Model Fits

Following Minda and Smith (2002), for each subject, we fit the models to each of four main trial segments of learning. In each of these trial segments, there were 10 repetitions of each stimulus. As an index of fit for the training data, we calculated the sum of squared deviations (*SSD*) between the observed and predicted Category-A response probabilities for all stimuli. Likewise, we fitted the models to the transfer data by minimizing the *SSD* between the predicted and observed response probabilities of the 16 transfer stimuli. For both the training and transfer data, we also used maximum-likelihood as a criterion of fit, and this method led to an entirely parallel set of results. We report the results from the *SSD* measure to increase the comparability with Minda and Smith's (2002) results.

Multiplicative-prototype model. Figure 1A displays the training-data fit values for the MPM and the probability-matching version of the GCM. Averaged across the four trial segments, the MPM's fit index was .153, and the fit index of the probability-matching GCM was .174. A repeated measures analysis of variance (ANOVA) with trial segment and model as factors indicated a main effect of model, $F(1, 47) = 5.878$, $p < .05$. There was no interaction between trial segment and model, $F(3, 141) = 1.37$, $p > .05$. Thus, just as in Minda and Smith's (2002) analysis of the data, these results point to a superiority of the prototype model over the restricted GCM.

However, a substantially different pattern of results emerged when the response-scaling parameter, γ, was allowed to vary in the exemplar model. Figure 1B shows the training-data fits of the GCM and the MPM for the four trial segments. The exemplar model's *SSD* was .109 compared with .153 for the prototype model. An ANOVA with trial segment and model as factors indicated that the GCM fit the data significantly better than the MPM, $F(1, 47) = 16.608$, $p < .001$, with no trial segment by model interaction, $F(2, 141) = 0.475$, $p > .05$.

We also fit the models to the transfer data. As reported by Minda and Smith (2002), the mean *SSD* for the probability-matching GCM (.668) and the mean *SSD* for the multiplicative prototype model (.693) were virtually identical. However, as with the training data, the full version of the GCM with γ as a free parameter

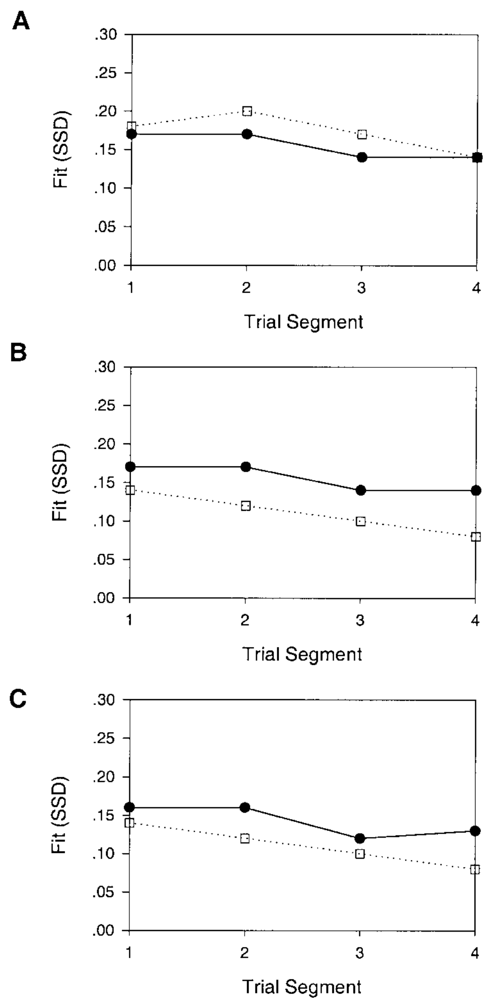

Figure 1. A: Fits of the multiplicative prototype model (MPM) and the probability-matching version of the generalized context model (GCM; with $\gamma = 1$) to each trial segment of the training phase in Experiment 2 of Minda and Smith (2002). Solid circles represent MPM; open squares represent GCM ($\gamma = 1$). B: Fits of the MPM and the full GCM to the same data. Solid circles represent MPM; open squares represent GCM. C: Fits of the mixed-prototype model (MIXED) and the GCM to the same data. SSD = sum of squared deviations. Solid circles represent MIXED; open squares represent GCM.

(mean $SSD = .337$) fit the transfer data significantly better than did the MPM (mean $SSD = .693$), $F(1, 47) = 19.747$, $p < .001$.

Two issues arise with regard to interpreting the superior fits of the exemplar model. First, in applying the prototype model to the Medin and Schaffer (1978) structure, we have followed previous approaches by assuming that the prototype of Category B is coded as 1 1 1 1. We refer to this approach as the *standard-coding assumption.* Note, however, that for Dimension 3, the values 0 and 1 occur an equal number of times in Category B (see Table 1). The standard-coding assumption is usually made to develop a contrast between the Category-A and Category-B prototypes. A question that arises is whether or not some subset of subjects may nevertheless develop an alternative representation in which the B prototype is coded as 1 1 0 1 (so that there is no contrast between the A and B prototypes along Dimension 3). It is straightforward to prove, however, that, for the present paradigm, the predictions of this alternative-coding prototype model can always be matched by a version of the standard-coding prototype model in which the Dimension-3 attention weight is set at zero. Thus, for the present paradigm, the alternative-coding prototype model can never provide fits to the data that are any better than those provided by the standard-coding model.

A second problematic issue is that the full version of the GCM has an extra free parameter compared with the prototype model and may achieve its better fits simply because it is a more flexible model (Minda & Smith, 2001). We took two approaches to addressing this concern. Although a detailed presentation goes beyond the scope of this article, in the first approach we conducted a variety of power analyses involving the models. Specifically, we generated simulated data sets under the assumption that the prototype model is the true model, and then fitted the GCM and the prototype model to these simulated data. In this case, the prototype model virtually always provided better fits than the GCM. In other words, although the GCM may be more flexible than the prototype model, we can at least conclude that if the prototype model were the true model underlying subjects' performance, then the GCM should not have provided better fits to the empirical data.

Our second approach to addressing this issue was to investigate the mixed-prototype model, which provides the prototype model with additional free parameters that have been deemed to be important in previous work (e.g., Smith & Minda, 1998; Smith et al., 1997). We report the results of these investigations in the following section.

Mixed-prototype model. In previous tests (Smith & Minda, 2000), the additional free parameters accorded the mixed-prototype model have been instrumental in allowing prototype theory to account for results from the Medin and Schaffer (1978) 5/4 paradigm. In addition, as noted previously, Smith and Minda (1998, 2000; Smith et al., 1997) have often acknowledged a role of exemplar memorization in situations involving extended training on a small set of exemplars. Therefore, we felt that it was important to test how well the mixed-prototype model could fit the data. Figure 1C shows the training-data fit of the mixed model compared with the full version of the GCM. Whereas the GCM's fit index was .109 across the four trial segments, the fit index for the mixed model was .141. An ANOVA with trial segment and model as factors indicated that the GCM fit the data significantly better than the mixed model, $F(1, 47) = 11.199$, $p < .01$. Once again, there was no trial segment by model interaction, $F(2, 141) = 0.602$, $p > .05$. In addition, the GCM fit the transfer data significantly better than did the mixed model (mean $SSD = .523$), $F(1, 47) = 10.520$, $p < .01$.

Locus of the effects. Although we fitted the models to the individual subject data, to gain some sense of the overall trends, we formed composite plots of the observed and predicted Category-A response probabilities for each stimulus. These composite plots, shown in Figure 2, were formed by averaging across the observed and predicted values for each individual observer. In general, the GCM captures the trends in the data better than the prototype or mixed model. The main qualitative advantage of the GCM is its ability to reproduce, in both the training and transfer-data sets, the observed advantage of Stimulus A2 over Stimulus A1. Minda and Smith (2002) reported that this advantage was

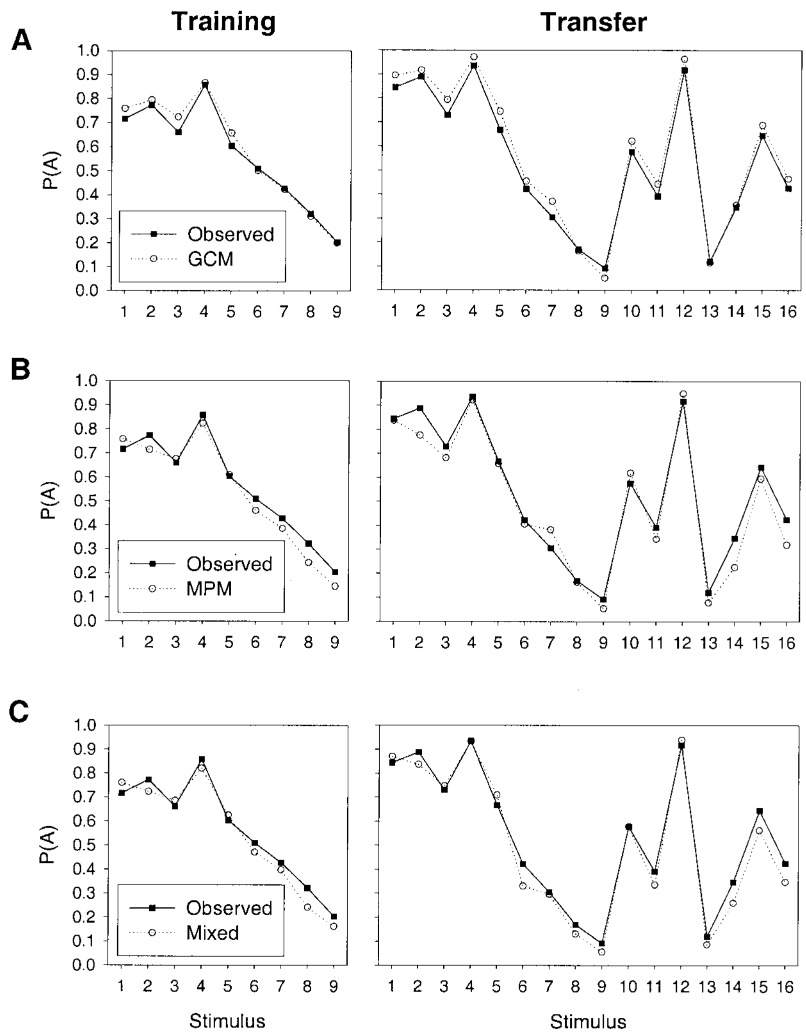

Figure 2. Averaged observed and predicted Category A response probabilities (P) for each individual stimulus in Experiment 2 of Minda and Smith (2002). The predicted values were obtained by averaging across fits to individual subject data. A: Predicted values are from the generalized context model (GCM). B: Predicted values are from the multiplicative prototype model (MPM). C: Predicted values are from the mixed-prototype model (Mixed).

statistically significant in the training data of both experiments, although it was not significant in the transfer data.

Finally, an examination of the best-fitting parameters from the GCM indicated that, not surprisingly, the γ response-scaling estimates tended to be greater than one. For example, in the fits to the transfer data, 41 of 48 subjects had γ estimates greater than one, with the median estimate being 4.86. This result is in accord with numerous previous ones indicating that individual subjects respond more deterministically than is predicted by the probability-matching version of the GCM.

In sum, when its response-scaling parameter is allowed to vary, the GCM provides a substantially better account than does the MPM or mixed model of the individual-subject data from Minda and Smith's (2002) replication of the Medin and Schaffer (1978) paradigm.

Attention Allocation

Background

The exemplar-model prediction of the A2-versus-A1 advantage depends, of course, on the assumption that subjects give at least some attention to Dimension 3 in making their classification judgments. (Otherwise, Exemplars A1 and A2 are logically equivalent, so there could be no basis for predicting differential performance on these items.) Minda and Smith (2002, p. 277), however, argued that it is more likely that subjects would ignore this dimension than grant it any attention because it is relatively nondiagnostic of category membership. Therefore, Minda and Smith (2002) suggested that the attention-weight estimates associated with Dimension 3 may have involved a formal artifice. This artifice allowed the GCM to capture the A2 advantage, but provided an incorrect psychological description of observers' actual attention-allocation behavior.

To corroborate this interpretation, Minda and Smith (2002) conducted two main attention-allocation analyses (described below). They argued that the results of these analyses confirmed that observers did not attend to the dimensions in the manner suggested by the exemplar model. Instead, these attention-allocation analyses supported the description provided by the prototype model.

In our view, Minda and Smith's (2002) attention-allocation analyses make potentially important contributions by seeking converging forms of evidence regarding the adequacy of formal psychological models of classification behavior. However, we believe that the particular analyses reported by Minda and Smith (2002) had some significant shortcomings. Contrary to Minda and Smith (2002), we argue in this section that their attention-allocation analyses do not distinguish between the predictions from exemplar and prototype models of classification.

Critique of Analysis 1

The first attention-allocation analysis conducted by Minda and Smith (2002) was as follows. Across the nine training stimuli, they simply computed the correlation between the values on each individual dimension and the observed Category-A response probabilities of each observer. For example, suppose that an observer tended to classify stimuli with value 0 on Dimension 1 into Category A and to classify stimuli with value 1 on Dimension 1 into Category B. Then there would be a large correlation computed for Dimension 1.

Minda and Smith (2002) interpreted these correlations as measures of attention to each individual dimension. They formed an overall correlation profile by averaging the correlations computed for each dimension across observers. They then compared this correlation profile with the average attention-weight distributions computed from the fitted models.

Examples of the results of these analyses are reproduced here as Figures 3A and 3B. These examples are from a subset of subjects in Minda and Smith's (2002) Experiment 1 who displayed a large A2-versus-A1 advantage. As can be seen in Figure 3A, the computed correlation along Dimension 3 is near 0, but the exemplar model assigns moderately large attention weight to this dimension in fitting the data. By contrast, the computed individual-dimension correlations tend to correspond well with the attention-weight estimates derived from the prototype model (Figure 3B). Thus, Minda and Smith (2002) argued that this attention-allocation analysis supported the prototype model and challenged the exemplar model.

Our view, however, is that this form of analysis is seriously flawed and has no bearing on the adequacy of the attention-allocation descriptions provided by the models. Although Minda and Smith (2002) advanced the correlation method as a model-free assay of attention, the method does entail explicit representational assumptions. To preview, a key problem is that the method assesses only independent relations between individual dimension values and classification performance and is insensitive to configural information in the category structure.

The most straightforward way to demonstrate the problem is by way of a simple example. Consider the category training structure illustrated in Table 2, which was used in a previous study reported by Medin, Altom, Edelson, and Freko (1982). The key aspect of the structure is that a perfect biconditional rule defined along Dimensions 3 and 4 is available for classifying the stimuli. That is, values of 11 or 00 along these dimensions signify Category A membership, whereas values of 10 or 01 along these dimensions signify Category B membership. Now, suppose that an observer learned this biconditional rule, applied it perfectly, and allocated all of his or her attention equally to these two dimensions (while ignoring Dimensions 1 and 2). Then the observer would classify training exemplars A1–A4 into Category A with probability one, and would classify training exemplars B1–B4 into Category B with probability one. The results from Minda and Smith's (2002) correlation-based analysis for this hypothetical example are pictured in Figure 3C. In a nutshell, their analysis would indicate that the observer gave *zero* attention to Dimensions 3 and 4 and gave *all* of his or her attention to Dimensions 1 and 2. In other words, Minda and Smith's (2002) correlation-based analysis would provide exactly the opposite picture of what was the true attention-allocation policy of the observer.

The reason for this misleading result is that Minda and Smith's (2002) correlation-based analysis treats each dimension independently. It is insensitive to any higher order configural relations among the dimensions. When considered independently, Dimensions 3 and 4 of the biconditional rule are uncorrelated with category membership. However, when considered configurally, they are perfectly correlated with category membership. The attention estimates provided by this correlation method tend to be similar to those obtained from the prototype model because the latter model also treats dimensions independently (Estes, 1986; Medin & Schaffer, 1978). By contrast, the exemplar model is sensitive to configural information across dimensions. Thus, its attention-weight estimates may often differ dramatically from those derived from Minda and Smith's (2002) correlation-based analysis.

The preceding example was intended to provide a simple illustration of the shortcomings of Minda and Smith's (2002) correlation method of assessing attention. Essentially the same problem exists, however, when the method is applied to the Medin and Schaffer (1978) 5/4 category structure. To demonstrate this point, we conducted the following analysis. First, we used the GCM to predict the Category-A response probabilities for the nine training stimuli in the 5/4 structure. In generating these predictions, we set the sensitivity parameter at $c = 4$, the response-scaling parameter at $\gamma = 3$, and assumed that the observer allocated equal attention to each of the four dimensions (i.e., $w_m = .25$ for each m). (Similar results are observed for a wide range of values of c and γ.) Next, we conducted Minda and Smith's (2002) correlation analysis on these GCM-predicted data. The results are shown in Figure 3D. As is evident from inspecting the figure, the correlation analysis suggests incorrectly that the observer gave nearly zero attention to Dimension 3, despite the fact that equal attention was in fact given to each of the four dimensions. In other words, if the GCM were the true model underlying observers' performance, then Minda and Smith's (2002) correlation analysis would incorrectly assess the degree of attention that observers gave to each dimension.

Minda and Smith (2002) acknowledged the potential limitations of their correlation method for assessing attention and wrote ". . . it is possible that the correlations may not always reflect participants' attentional allocation" (p. 283). Neverthe-

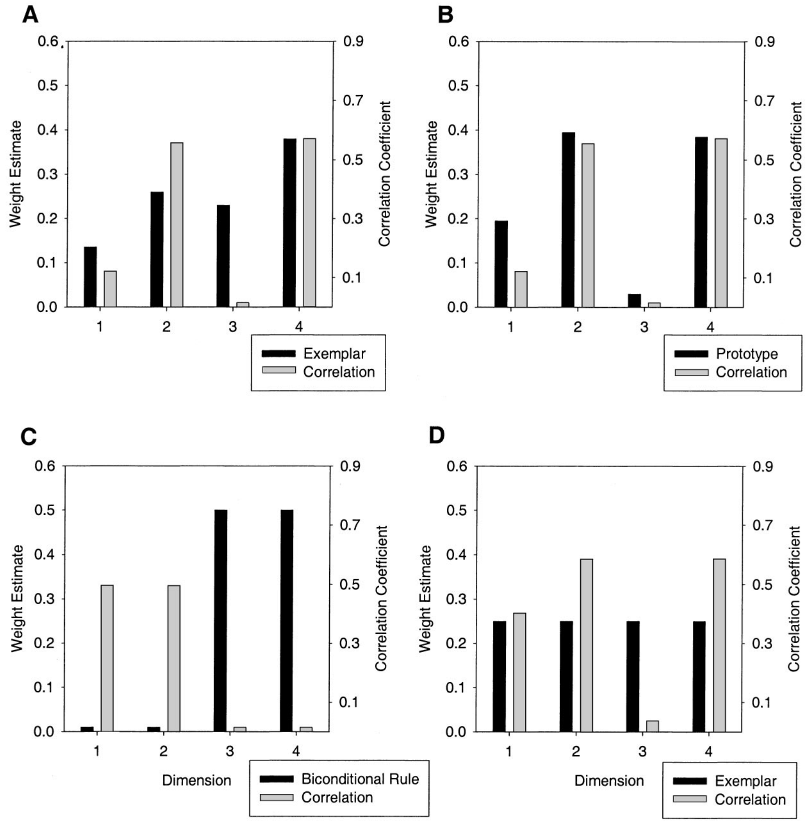

Figure 3. A: Dark bars are the average attention weights estimated by the exemplar model for each of the four dimensions for the subjects with a large A2 advantage in Experiment 1 of Minda and Smith (2002). Light bars are the average correlation coefficients of these subjects' responses to the nine stimuli with the binary-valued features of the stimuli (shown separately for each of the four dimensions). B: Dark bars are the average attention weights for each of the four dimensions estimated by the prototype model for the same subjects. Light bars are the average correlation coefficients of their responses with the featural values of the stimuli. C: Dark bars are the attention weights of a hypothetical observer who placed all attention on Dimensions 3 and 4 in Medin, Altom, Edelson, and Freko's (1982) Experiment 4 design. The light bars are the results from Minda and Smith's (2002) correlation-based analysis for this hypothetical example. Zero values were plotted at .01 in this panel for clarity. D: Dark bars are the generalized context model attention weights of a hypothetical observer who distributed his or her attention equally among the four dimensions in the Medin and Schaffer (1978) design. The light bars are the results from Minda and Smith's (2002) correlation-based analysis for this hypothetical example.

less, they repeatedly included the results of these analyses as evidence bearing on the adequacy of the alternative models (e.g., Minda & Smith, 2002, pp. 282, 286, 287, 288). The preceding demonstrations show clearly that Minda and Smith's (2002) concerns about this mode of analysis were highly justified. In a nutshell, this first method of assessing attention allocation has no bearing on evaluating prototype versus exemplar models of classification.

Critique of Analysis 2

To address the aforementioned concerns, Minda and Smith (2002) reported results from a second, more intricate method of assessing attention allocation in the formal models. In this method, they simulated results from the models themselves to assess the expected correlations between the dimension values and the classification response probabilities of the nine training stimuli. They

Table 2
Medin et al.'s (1982) Experiment 4 Category Structure

	Dimension			
Stimulus	1	2	3	4
Category A				
A1	1	1	1	1
A2	1	1	0	0
A3	0	1	1	1
A4	1	0	0	0
Category B				
B1	0	0	1	0
B2	0	0	0	1
B3	1	0	1	0
B4	0	1	0	1

then analyzed whether the results of fitting the models to the observed classification data conformed to these expected correlations.

Specifically, Minda and Smith's (2002) second mode of analysis proceeded as follows. They started by simulating 5,000 subjects from each of the models. For each simulated subject, they chose a random configuration of the models' parameters, and for each of these parameter configurations they computed the predicted Category-A response probabilities for each of the stimuli in the Medin and Schaffer (1978) 5/4 training set. Next, for each of the simulated subjects, they calculated a performance-structure correlation by correlating these predicted probabilities with the logical feature values of the stimuli along a particular dimension. They then plotted the performance-structure correlation against the attention-weight parameter value that was used to generate the predicted probabilities for that simulated subject. By plotting these 5,000 points, Minda and Smith (2002) constructed a response surface for the models in which all possible performance-structure correlations for particular settings of the attention-weight parameters were mapped out.

Minda and Smith's (2002) next step was to compare the observed performance-structure correlations with the ones predicted from the models. Thus, across the nine training stimuli, they calculated for each subject in each of the last two trial segments of training the correlation between the observed Category-A response probabilities and the abstract feature values along a particular dimension. They also fitted each model to the individual subject's classification data to estimate the best-fitting attention weights along each dimension. Finally, they plotted the observed performance-structure correlations against these best-fitting attention weights. The goal was to discover whether the observed performance-structure correlations fell within the range that was predicted by the models in the response-surface analysis.

Minda and Smith (2002) plotted the Dimension 1 and Dimension 3 performance-structure correlations for both the GCM (with $\gamma = 1$) and the MPM. These plots are shown here in Figures 4 and 5, respectively. For Dimension 1, for both models, most of the observed performance-structure correlations lie within the range that is predicted by the models. However, for Dimension 3, there is a different pattern of results for the GCM. In this case, the observed performance-structure correlations often lie outside of the predicted range. On the basis of this analysis, Minda and Smith (2002) argued that the GCM therefore provided an incorrect description of individual subjects' attention-allocation behavior,

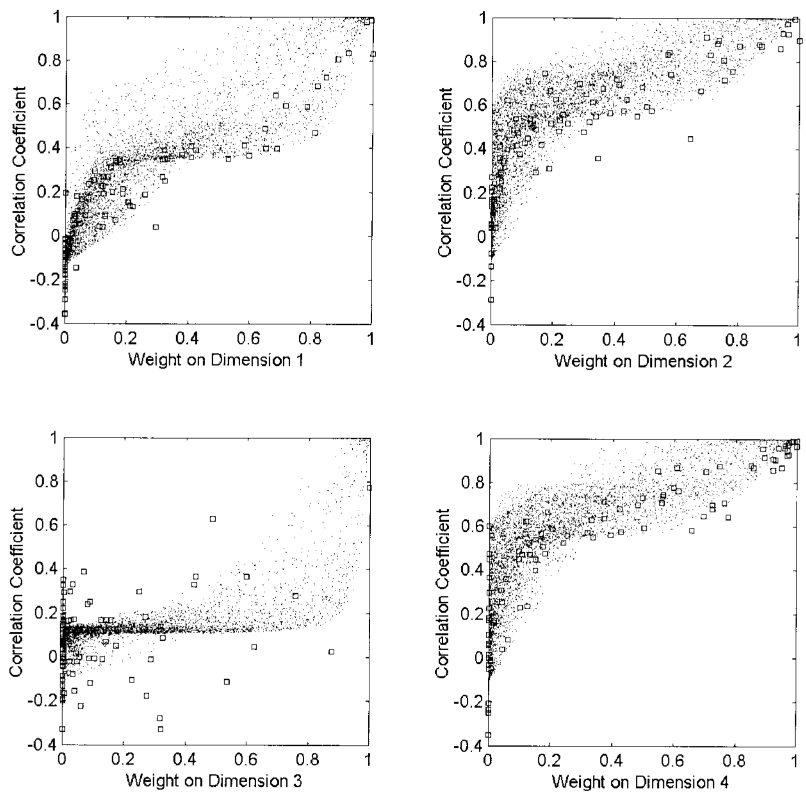

Figure 4. Performance-structure correlations scatterplots for the generalized context model (with $\gamma = 1$) for each of the four dimensions. The *x*-axis represents the weights that were assigned to that dimension in the model. The *y*-axis represents the correlation between performance on the training stimuli and their featural values on that dimension. Each small dot is a performance-structure correlation coefficient plotted against the attention-weight parameter value that was used to generate the predicted probabilities for one of the 5,000 simulated subjects. The small squares represent the performance-structure correlations observed in Experiment 2 of Minda and Smith (2002) plotted against the attention weight estimated by the model.

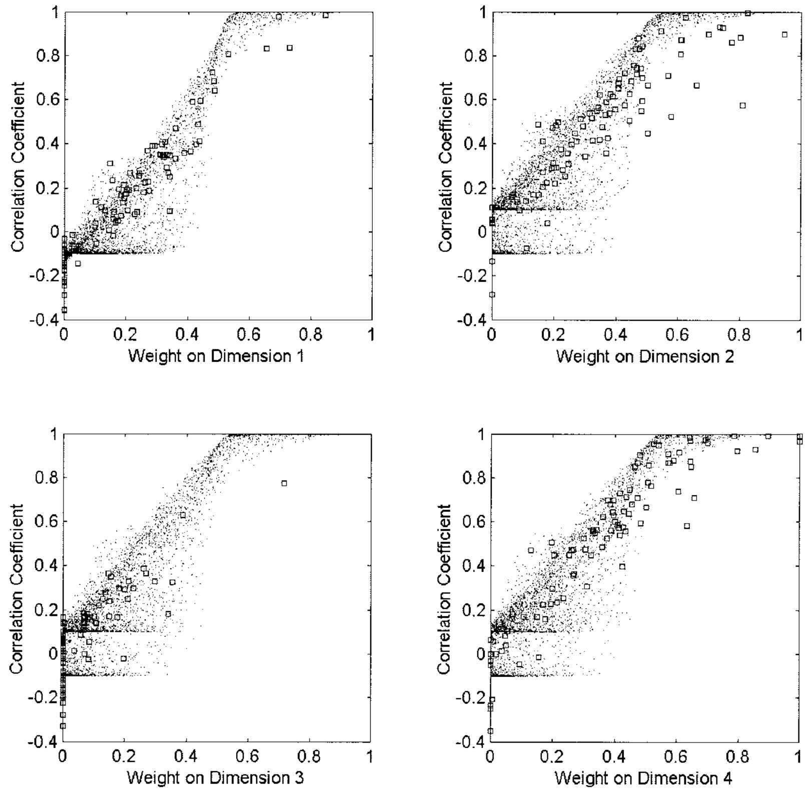

Figure 5. Performance-structure correlations scatterplots for the multiplicative prototype model for each of the four dimensions. The *x*-axis represents the weights that were assigned to that dimension in the model. The *y*-axis represents the correlation between performance on the training stimuli and their featural values on that dimension. Each dot is a performance-structure correlation coefficient plotted against the attention-weight parameter value that was used to generate the predicted probabilities for one of the 5,000 simulated subjects. The squares represent the performance-structure correlations observed in Experiment 2 of Minda and Smith (2002) plotted against the attention weight estimated by the model.

whereas the description provided by the prototype model was supported.

Although this second mode of analysis does not have the same problems as does the first, our view is that some serious shortcomings remain. First, in conducting the analysis, Minda and Smith (2002) examined the attention-allocation plots for only two of the dimensions (1 and 3). In Figures 4 and 5 we provide a complete analysis, which also shows the performance-correlation plots for Dimensions 2 and 4. As it turns out, the prototype model does poorly on these dimensions, whereas the GCM does well. That is, although the GCM (with $\gamma = 1$) fares poorly on the Dimension-3 analysis, the prototype model fares poorly on Dimensions 2 and 4. We should note that this shortcoming of the prototype model was not as evident for Minda and Smith's (2002) Experiment 1 training data. Nevertheless, the point is that there are cases in which the prototype model too does poorly given this mode of analysis.

A more important problem with Minda and Smith's (2002) analysis is that it made no allowance for sampling error in deriving the expected range of performance-structure correlations. The derived response surfaces are population-based predictions, that is, they are based on an infinite number of responses given the parameter settings in the models. However, in the actual analysis of the experimental data, subjects contributed only 10 observations for each individual stimulus. To investigate the effect of sampling error on this analysis, we used the following procedure. First, for each of 5,000 random parameter settings, we generated the classification probabilities predicted from the models. Given these predicted probabilities, we then simulated 10 binary-valued responses for each of the training stimuli (assuming binomial variability). The simulated Category-A response probability for that stimulus was then found by averaging across the results of these 10 simulated responses. We applied this procedure for each of the stimuli for each of our 5,000 simulated subjects. Then, as before, for each subject, we calculated a performance-structure correlation by correlating the simulated probabilities with the logical feature values of the stimuli along a particular dimension. The resulting composite plots yield the models' predicted range of performance-structure correlations, given sampling error. The plots for the GCM (with $\gamma = 1$) and the MPM are shown in Figures 6 and 7, respectively. For both of the models, the observed performance-structure correlations generally fall within the

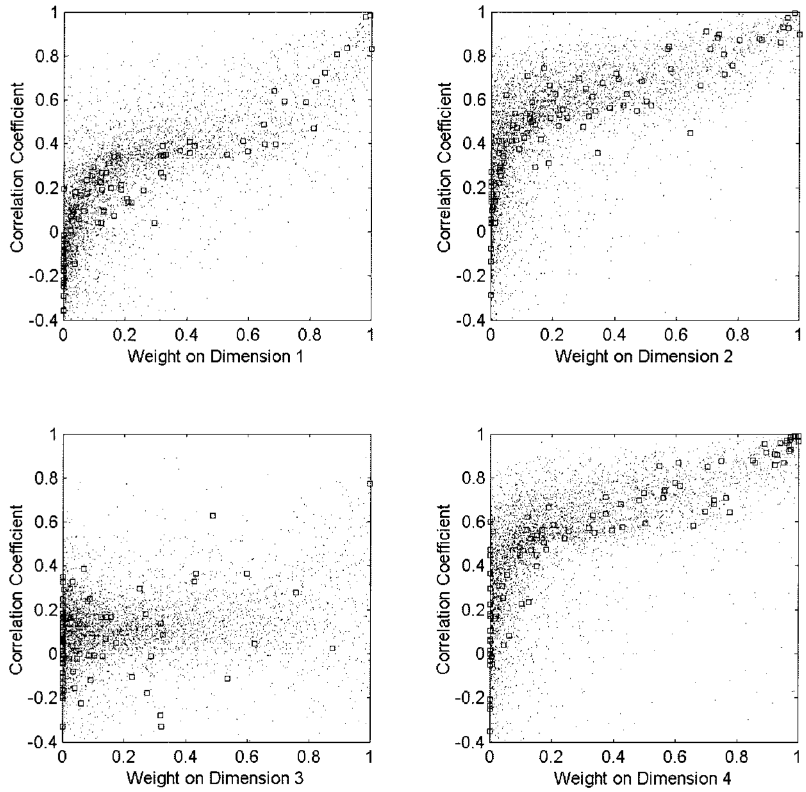

Figure 6. Performance-structure correlations scatterplots, given sampling error, for the generalized context model (with $\gamma = 1$) for each of the four dimensions. The *x*-axis represents the weights that were assigned to that dimension in the model. The *y*-axis represents the correlation between performance on the training stimuli and their featural values on that dimension. Each dot is a performance-structure correlation coefficient plotted against the attention-weight parameter value that was used to generate the predicted probabilities for one of the 5,000 simulated subjects. The squares represent the performance-structure correlations observed in Experiment 2 of Minda and Smith (2002) plotted against the attention weight estimated by the model.

range of the predicted performance-structure correlations. Apparently, when allowance is made for sampling error, Minda and Smith's (2002) correlation-based analysis does not distinguish between the models in this paradigm.[2]

Finally, as in their other modeling analyses, Minda and Smith (2002) considered only the version of the GCM that assumes a probability-matching response rule. We were interested, therefore, in investigating the predictions from the full version of the GCM that makes allowance for response-scaling processes. In particular, we conducted the same performance-structure correlation analysis as did Minda and Smith (2002), except now also allowing γ to vary randomly between 0 and 20. As can be seen in Figure 8, for the full version of the GCM, even without making allowance for sampling error, the observed performance-structure correlations fall within expected bounds given the range of parameter settings of this model. When allowance is made for sampling error, the range of expected performance-structure correlations expands even more. We conclude that Minda and Smith's (2002) analysis does not pose a challenge to the GCM's description of attention allocation in classification.

General Discussion

Model Comparisons

One of the classic paradigms for distinguishing between prototype and exemplar models of category representation is the Medin and Schaffer (1978) 5/4 structure. Within this structure, prototype models predict that Training Exemplar A1 will be classified with higher accuracy than will Training Exemplar A2, whereas exemplar models generally predict the opposite. Across the vast majority of studies in which the actual 5/4 structure was tested, the qualitative prediction from exemplar models has been supported and the prediction from prototype models disconfirmed (Nosofsky, 2000). Furthermore, the quantitative fits of a well-known exemplar model, the GCM (Medin & Schaffer, 1978; Nosofsky, 1986), are superior to those of prototype models when these models are fitted to the complete sets of classification-transfer data observed in

[2] One potential criticism of the GCM is that for Dimension 3, the model is compatible with a very large range of performance-structure correlations. We are not claiming that these results demonstrate a success of the model, but rather that they do not pose a challenge for the model.

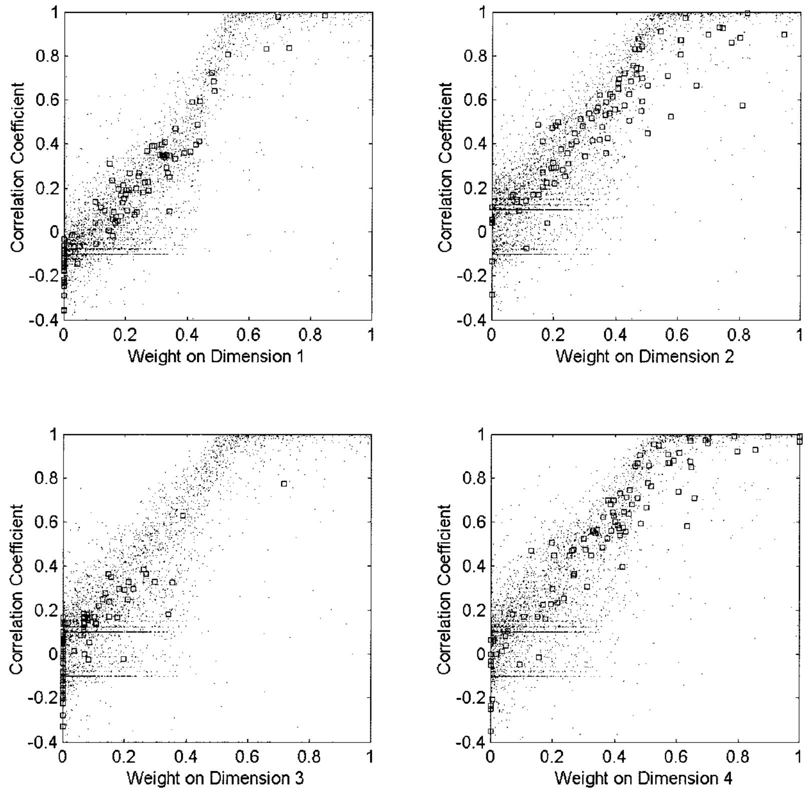

Figure 7. Performance-structure correlations scatterplots, given sampling error, for the multiplicative prototype model for each of the four dimensions. The *x*-axis represents the weights that were assigned to that dimension in the model. The *y*-axis represents the correlation between performance on the training stimuli and their featural values on that dimension. The dots are the performance-structure correlation plotted against the attention-weight parameter value that was used to generate the predicted probabilities (with sampling error) for each of 5,000 simulated subjects. The squares represent the performance-structure correlations observed in Experiment 2 of Minda and Smith (2002) plotted against the attention weight estimated by the model.

these numerous experiments (Nosofsky, 1992; Smith & Minda, 2000).

As aptly pointed out by Minda and Smith (2002), however, previous tests compared the quantitative fits of these models with data averaged across subjects. There is clear evidence, however, that there are large individual differences in patterns of classification in this paradigm (e.g., Nosofsky, Palmeri, & McKinley, 1994). Therefore, the averaged data are not representative of patterns of performance at the individual-observer level. Furthermore, other researchers have suggested that the context model may be advantaged relative to alternative models in fitting averaged data (e.g., Maddox, 1999).

Thus, Minda and Smith (2002) took an important step in considering the ability of exemplar and prototype models to fit individual-subject data in the Medin and Schaffer (1978) 5/4 paradigm. On the surface, their finding that the prototype model actually provided better quantitative fits to the individual-subject data than did the context model suggests that the previous support for exemplar models may indeed have involved certain artifacts due to averaging across subjects. On the other hand, Minda and Smith's (2002) finding of a consistent advantage for Stimulus A2 over Stimulus A1 across two experiments and in both learning and transfer is at odds with a prototype-based account.

Our view is that Minda and Smith's (2002) model comparisons were limited in that they considered only a restricted version of the exemplar model. Specifically, whereas the original version of the context model assumed a probability-matching response rule, it is well known that individual subjects respond more deterministically than is predicted by a probability-matching strategy. Plausible applications of the context model to fitting individual-subject data require the use of an extended version of that model that makes allowance for response-scaling processes (Ashby & Maddox, 1993; McKinley & Nosofsky, 1995; Nosofsky, 1991). Furthermore, as argued by Nosofsky and Zaki (2002), such a response-scaling mechanism can be viewed as implicit in the prototype model that Minda and Smith (2002) tested. Thus, Minda and Smith's (2002) finding of better fits of the prototype model compared with the context model may have nothing to do with the alternative representational assumptions of these models (i.e., prototypes vs. exemplars). Rather, they may simply reflect that the version of the prototype model they tested could accommodate the deterministic response strategies exhibited by individual subjects, whereas the probability-matching version of the context model could not. Accordingly, one of the central goals of the present research was to fit the current version of the exemplar model that

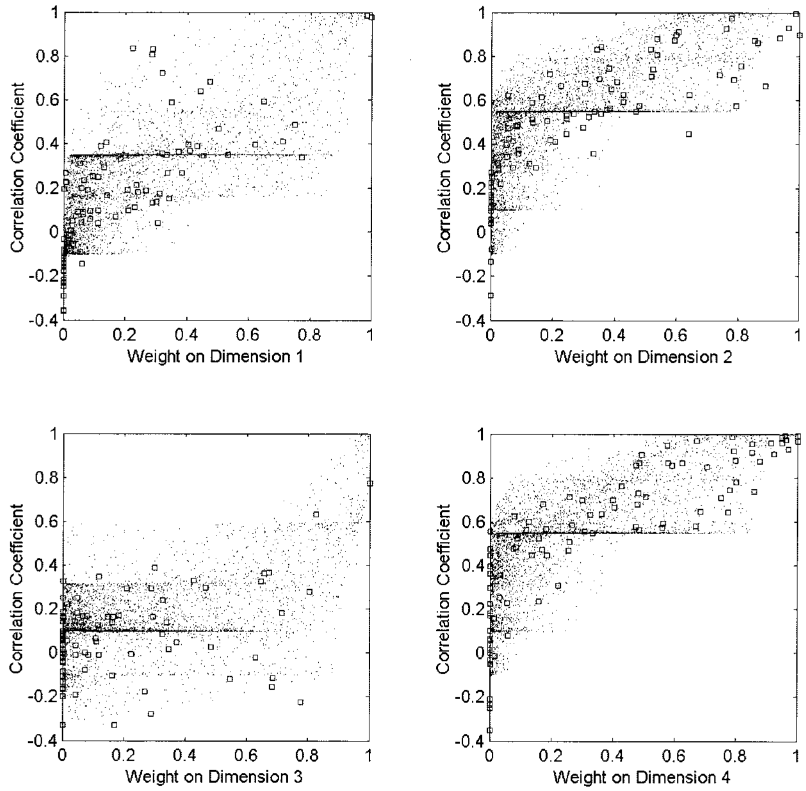

Figure 8. Performance-structure correlations scatterplots for the generalized context model for each of the four dimensions. The *x*-axis represents the weights that were assigned to that dimension in the model. The *y*-axis represents the correlation between performance on the training stimuli and their featural values on that dimension. The small dots are the performance-structure correlation plotted against the attention-weight parameter value that was used to generate the predicted probabilities for each of 5,000 simulated subjects. The small squares represent the performance-structure correlations observed in Experiment 2 of Minda and Smith (2002) plotted against the attention weight estimated by the model.

makes allowance for response-scaling processes to Minda and Smith's (2002) individual-subject data in the 5/4 paradigm.

The results of our modeling analyses of Minda and Smith's (2002) data were clear-cut. The full version of the exemplar model provided significantly better fits to the individual subject data in the 5/4 paradigm than did any of the competing models. Furthermore, unlike the prototype and mixed models, it accounted for the A2-versus-A1 advantage that was so consistently observed in Minda and Smith's (2002) data.

Minda and Smith (2002, p. 290) acknowledged the inability of the prototype model to account for the A2 advantage and attributed the result to some unspecified "secondary" process or representational system. One type of secondary system that they have posited in past work is the all-or-none exemplar-memorization process that forms part of their proposed mixed-prototype model. As discussed in previous works, however, this particular secondary form of category representation fails to predict the A2 advantage (Nosofsky, 2000; Smith & Minda, 2000). And as shown in the present research, the exemplar model provides substantially better quantitative fits to Minda and Smith's (2002) individual-subject data than does this mixed model.

Our view is that a more plausible version of a mixed prototype-plus-exemplar model would make allowance for the idea that the stored exemplars can also support generalization decisions. Given that such an exemplar-based generalization mechanism is already formalized in the GCM, however, the question then becomes whether there is a need to extend the model with a prototype-based mechanism as well. To investigate this issue, in the Appendix we report additional model-based analyses of Minda and Smith's (2002) individual-subject data. Specifically, we test the predictions of a mixed prototype-plus-exemplar model in which classification decisions are based on similarity comparisons to both stored exemplars and a prototype. Because the GCM arises as a special case of this mixed model in which the prototype representation is given zero weight, the GCM cannot provide a better absolute fit to the data than the mixed model. Instead, the question of interest is whether the mixed model provides a substantial improvement compared with the GCM in its ability to account for the classification data. In a nutshell, the results of our model comparisons provided rather little evidence of the utility of extending the exemplar model with a prototype-abstraction mechanism (see the Appendix for details). It is an open question whether alternative versions of a mixed prototype-plus-exemplar model would yield more convincing evidence of the operation of prototypes.

Attention Allocation

The other major theme of Minda and Smith's (2002) study was the argument that the best-fitting version of the GCM provided an untenable description of individual subject's attention allocation behavior. In the present article, however, we pointed to important shortcomings in Minda and Smith's (2002) attention-allocation analyses. We showed that their correlation-based method of assessing attention was flawed. For example, we showed that if the context model were the true model that governed performance, then Minda and Smith's (2002) correlation-based method would not recover the true attention weights that observers give to the component dimensions of the stimuli. In addition, we showed that the more sophisticated response–surface version of their analysis was also flawed. First, in some cases they did not report aspects of the analysis that challenged the prototype model itself. Second, their analysis made no allowance for sampling error in assessing the range of the models' predictions. Third, their analysis focused on only the probability-matching version of the GCM, which has already been acknowledged to be inadequate in modeling individual-subject behavior. When these shortcomings were corrected, our analysis found no evidence that challenged the exemplar model's description of attention allocation in the 5/4 paradigm.

In conclusion, contrary to Minda and Smith's (2002) claims, we argue that the results reported by Minda and Smith (2002) do not challenge the assumptions of the GCM. In fact, the model-fitting comparisons on the individual-subject data, as well as the consistent A2-versus-A1 advantage, strongly favor the exemplar model over the prototype and mixed models. We do not argue here that the exemplar model is the "correct" model, only that it remains as providing a viable account of performance in this paradigm.

References

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33–53.

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

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155–174.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception and Psychophysics, 61,* 354–375.

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

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, and Cognition, 27,* 775–799.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 275–292.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and 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, and Cognition, 13,* 87–108.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.

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

Nosofsky, R. M. (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., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7,* 375–402.

Nosofsky, R. M. & Palmieri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Nosofsky, R. M., Palmieri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

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, and Cognition, 28,* 924–940.

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.

Shepard, R. N. (1987, September 11). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

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, and Cognition, 24,* 1411–1436.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 3–27.

Smith, J. D., Murray, M. J., Jr., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 659–680.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44,* 92–107.

Appendix

Tests of a Weighted Prototype-Plus-Exemplar Model

In this appendix we report results of analyses involving a weighted prototype-plus-exemplar (WPE) model. According to the WPE model, people represent categories by storing both exemplars and prototypes in memory and classify objects on the basis of how similar they are to both forms of category representation. The WPE generalizes both the MPM and the GCM, that is, the latter two models arise as special cases of the WPE with appropriate restrictions of its free parameters.

According to the WPE model, the evidence in favor of Category K, given presentation of item i, $E(i, K)$, is given by

$$E(i, K) = v * Exemp(i, K) + (1 - v) * Prot(i, K), \tag{A1}$$

where $Exemp(i, K)$ is the summed similarity of item i to the exemplars of Category K (as computed in Equations 1 and 2); $Prot(i, K)$ is the similarity of item i to the prototype of Category K (as computed in Equations 4 and 5); and v ($0 \leq v \leq 1$) is a representation-weight parameter. The probability with which item i is classified into Category A is given by

$$P(A/i) = \frac{E(i, A)^{\gamma}}{E(i, A)^{\gamma} + W(i, B)^{\gamma}}, \tag{A2}$$

where γ is the response-scaling parameter. When $v = 1$ in Equation A1, the model reduces to the GCM; whereas when $v = 0$ the model reduces to the MPM. We assume that the same sensitivity and attention-weight parameters operate in the computation of exemplar-based and prototype-based similarity. Thus, the free parameters in the WPE model are overall sensitivity c, the response-scaling parameter γ, the attention weights w_m, and the representation-weight parameter v. The WPE model was introduced in previous work conducted by Shin and Nosofsky (1992), although the earlier version did not include the γ response-scaling parameter.

We compared the WPE model to the GCM by using the Bayesian information criterion (BIC) statistic (e.g., Wasserman, 2000). The BIC statistic is given by

$$\mathrm{BIC} = -2 * (\log \mathrm{L}) + \mathrm{d} * \log(\mathrm{N}), \tag{A3}$$

where log L is the (natural) log-likelihood of the data given the model, d is the number of free parameters in the model, and N is the number of data observations on which the model fit is based. The model that yields the smaller value of the BIC statistic is preferred. Note that the BIC statistic penalizes a model for its number of free parameters. Thus, if a simpler model with fewer free parameters yields essentially the same log-likelihood fit as does a more complex model, then the simpler model will be preferred.

We fitted both the WPE model and the GCM to the individual-subject transfer data collected by Minda and Smith (2002) by searching for the free parameters that minimized the BIC statistic. Among the 48 data sets, the WPE model yielded a smaller BIC value than did the GCM in only six cases. The mean BIC yielded by the GCM (59.1) was significantly smaller than the mean BIC yielded by the WPE model (61.5), $t(47) = 3.31$ $p < .01$. We interpret these results as providing rather weak evidence for the utility of extending the GCM with a prototype-based representation.

Received June 4, 2002
Revision received March 15, 2003
Accepted March 17, 2003 ■

Correction to Oberauer and Wilhelm (2003)

The article “The Meaning(s) of Conditionals: Conditional Probabilities, Mental Models, and Personal Utilities,” by Klaus Oberauer and Oliver Wilhelm (*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2003, Vol. 29, No. 4, pp. 680–693), is corrected as follows: On page 684, Table 4, all correlations should have been identified as having a $p < .05$. On page 689, Table 7, the ratio index in column 2 and the frequency index in row 1 were indicated as ≥ 40. They should have appeared as ≤ 40.