Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 5, 924–940

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.5.924

Exemplar and Prototype Models Revisited: Response Strategies, Selective Attention, and Stimulus Generalization

Robert M. Nosofsky and Safa R. Zaki
Indiana University Bloomington

J. D. Smith and colleagues (J. P. Minda & J. D. Smith, 2001; J. D. Smith & J. P. Minda, 1998, 2000; J. D. Smith, M. J. Murray, & J. P. Minda, 1997) presented evidence that they claimed challenged the predictions of exemplar models and that supported prototype models. In the authors' view, this evidence confounded the issue of the nature of the category representation with the type of response rule (probabilistic vs. deterministic) that was used. Also, their designs did not test whether the prototype models correctly predicted generalization performance. The present work demonstrates that an exemplar model that includes a response-scaling mechanism provides a natural account of all of Smith et al.'s experimental results. Furthermore, the exemplar model predicts classification performance better than the prototype models when novel transfer stimuli are included in the experimental designs.

A classic issue in cognitive psychology concerns the manner in which people represent categories in memory. According to prototype models (Homa, 1984; Posner & Keele, 1968; Reed, 1972), people represent categories by forming a summary representation that is a central tendency of all of the experienced members of a category. Classification decisions are based on the similarity of an item to the alternative prototypes. By contrast, according to exemplar models (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986), people represent categories by storing the individual members or exemplars of a category as separate traces and classify items based on their similarity to these stored exemplars.

There is a good deal of converging evidence that prototype models are insufficient as models of categorization, especially in situations in which learners have been repeatedly exposed to the individual exemplars of each category. Memories for the individual exemplars appear to play at least some role in such situations (Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968; Nosofsky, 1992; Smith & Minda, 2000).

Conversely, however, in a recent series of articles, Smith and colleagues (Minda & Smith, 2001; Smith & Minda, 1998, 2000; Smith, Murray, & Minda, 1997) have presented evidence that they claim severely challenges exemplar models of categorization. These researchers argue that the formation of prototypes serves as a first organizing principle in the representation of categories. They hypothesize that, especially at early stages of learning, people's category representations are prototype based. With continued learning on a restricted set of exemplars, people may eventually form memories for those exemplars and use those memories to classify the old training instances. However, generalization to new items is still assumed to be governed by similarity to the prototype.

In this article we question the evidence that Smith et al. (1997; Smith & Minda, 1998) have advanced in favor of the prototype hypothesis and against the exemplar view. Specifically, we argue that a current exemplar-based generalization model of categorization provides a natural account of all of their results. Furthermore, we present evidence that their alternative "mixed model" of categorization, which involves a combination of prototype-based representation together with all-or-none memories for specific exemplars, fails to predict correctly people's classification performance.

We organize our article by first presenting the main competing models of classification that serve as representatives of the prototype and exemplar views. Next, we review the key experimental paradigm and sources of evidence that Smith et al. have used to challenge the exemplar-generalization model and to support the prototype model. We then explain why we believe that Smith and colleagues' paradigm and results fail to discriminate between these alternative modeling approaches and suggest that their data provide little evidence for the operation of a broad-based prototype-abstraction process. Finally, we report a new series of experiments to corroborate our interpretations and to develop contrasts between the predictions from the exemplar-generalization and mixed-prototype models.

Review of the Models

In this section we review the various models that serve as representatives of the exemplar, prototype, and mixed-representation views. In the experimental paradigms under consideration, all of the models are compared in situations in which the stimuli are composed of binary-valued, separable dimensions and in which subjects are classifying the stimuli into one of two categories. For simplicity, we describe the models as they are applied in such a paradigm, although extensions of the models to more general paradigms are straightforward (e.g., Nosofsky, 1986, 1987; Shin & Nosofsky, 1992).

Robert M. Nosofsky and Safa R. Zaki, Department of Psychology, Indiana University Bloomington.

This work was supported by National Institute of Mental Health Grant R01 MH48494. We thank Michael Erickson and two anonymous reviewers for their helpful criticisms of an earlier version of this article.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Department of Psychology, Indiana University, Bloomington, Indiana 47405. E-mail: nosofsky@indiana.edu

Exemplar-Generalization Model

The representative of the class of exemplar-generalization models is the *context model* of classification proposed by Medin and Schaffer (1978), generalized by Nosofsky (1984, 1986), and further extended by Ashby and Maddox (1993). According to the model, the probability that item i is classified into Category A is given by

$$P(\mathrm{A}|i) = \frac{(\sum s_{\mathrm{ia}})^{\gamma}}{(\sum s_{\mathrm{ia}})^{\gamma} + (\sum s_{\mathrm{ib}})^{\gamma}}, \tag{1}$$

where $\Sigma\, s_{\mathrm{ia}}$ and $\Sigma\, s_{\mathrm{ib}}$ denote 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). The γ parameter, introduced into the context-model response rule by Ashby and Maddox (1993), governs the extent to which responding is probabilistic versus deterministic. When $\gamma = 1$, observers respond probabilistically by "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 larger summed similarity.

Issues pertaining to the γ response-scaling parameter are critical to evaluating the comparisons between exemplar and prototype models conducted by Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998). In all of the model comparisons that they conducted, these researchers set $\gamma = 1$ in the exemplar model. However, previous research has demonstrated that this constrained version of the exemplar model is inadequate because it fails to account for the deterministic patterns of responding that are often evidenced at the individual subject level (e.g., Ashby & Gott, 1988; Ashby & Maddox, 1993; Maddox & Ashby, 1993; McKinley & Nosofsky, 1995; Nosofsky, 1991a). Much of our discussion in the present article involves reviewing the importance of the γ response-scaling parameter and demonstrating that analogous forms of response scaling are just as important to prototype-based accounts of classification performance as to exemplar-based accounts.

According to the context model, in situations involving binary-valued, separable dimensions, the distance between exemplars i and j is given by

$$d_{\mathrm{ij}} = \sum_{m=1}^{M} w_{\mathrm{m}} \cdot |x_{\mathrm{im}} - x_{\mathrm{jm}}|, \tag{2}$$

where x_{im} and x_{jm} denote the values of exemplars i and j on dimension m, respectively, w_{m} is the "attention weight" given to dimension m, and M denotes the number of dimensions along which the stimuli vary. The attention-weight parameters are constrained to vary between 0 and 1 ($0 \le w_{\mathrm{m}} \le 1$) and to sum to 1. Note that because the dimensions are binary-valued, each component $|x_{\mathrm{im}} - x_{\mathrm{jm}}|$ is equal either to 0 (if exemplars i and j match on dimension m) or to 1 (if the exemplars mismatch on this dimension).

The selective attention process formalized in Equation 2 is an essential aspect of the exemplar-based context model. As discussed by Medin and Schaffer (1978) in their original formulation of the model, the selective attention process may capture the types of "hypothesis testing" behavior presumed to underlie classification performance, especially at early stages of learning. In addition, there is evidence that with extended training, observers often learn to distribute attention across dimensions in a manner that tends to optimize performance (e.g., Nosofsky, 1984, 1986, 1991a).

The similarity between exemplars i and j is an exponential decay function of their distance (Shepard, 1987),

$$s_{\mathrm{ij}} = \exp(-c \cdot d_{\mathrm{ij}}), \tag{3}$$

where c is an overall sensitivity parameter that determines the rate at which similarity declines with distance. As explained in numerous previous articles, the combination of the "city-block" distance metric formalized in Equation 2 and the exponential similarity function formalized in Equation 3 yields an interdimensional multiplicative rule for computing similarity (see Medin & Schaffer, 1978; and Nosofsky, 1984, for more extensive discussion). As a result, the context model is often referred to as a multiplicative-similarity exemplar model.

The similarities computed from Equations 2 and 3 are substituted into Equation 1 to yield the classification probabilities that are predicted by the model. The free parameters in the model are the response-scaling parameter γ, the sensitivity parameter c, and $M - 1$ freely varying attention weights (w_{m}).

Additive-Prototype Model

In the various prototype models under consideration, a prototype vector is defined for each of Categories A and B. The prototype for Category A is composed of the dimension values that occur most frequently among the members of Category A, and likewise for the prototype of Category B. According to the additive-prototype model advanced by Smith et al. (1997; Smith & Minda, 1998), when item i is presented, the evidence in favor of a Category A response is given by

$$\mathrm{E}_{\mathrm{i,A}} = \sum \delta_{\mathrm{m}} \cdot w_{\mathrm{m}}, \tag{4}$$

where w_{m} is the attention weight given to Dimension m (just as in the context model's Equation 2); and δ_{m} is an indicator variable set equal to 1 if item i matches the prototype of Category A on dimension m, and set equal to zero if it mismatches the prototype of Category A on this dimension. The probability that an observer classifies item i into Category A is given by

$$P(\mathrm{A}|i) = \frac{g}{2} + (1 - g) \cdot \mathrm{E}_{\mathrm{i,A}}, \tag{5}$$

where g ($0 \le g \le 1$) is a guessing parameter. That is, with probability g the observer guesses randomly between the two categories, and with probability $(1 - g)$ the observer uses the prototype-based evidence to make a response. The guessing parameter is critical to the additive-prototype model if it is to make plausible predictions of classification performance. If $g = 0$, then the item that is the prototype of each category is predicted to be classified with probability 1 into its correct category. The guessing parameter allows the additive-prototype model to account for the less-than-perfect performance that is sometimes observed for the prototype patterns.

Note that the additive-prototype model advanced by Smith et al. (1997; Smith & Minda, 1998) does not include a response-scaling parameter. In the section The Prototype-Plus-Exception Structure, we explain why this highly simplified model is often sufficient to account for performance at very early stages of classification learning. We go on to demonstrate in our subsequent experiments, however, that this highly simplified model is quite limited in generality. Even if the category representation is prototype based, we demonstrate that the additive-prototype model needs to be augmented with a response-scaling parameter, analogous to the one found in the context model.

Multiplicative-Prototype Model

Besides assuming a different category representation, the context model and the additive-prototype model differ in the rules that are used to compute "similarity," that is, the context model uses a multiplicative rule whereas the additive-prototype model uses an additive rule. To improve the comparability between the models, Estes (1986) and Nosofsky (1987, 1992) introduced a multiplicative version of the prototype model that uses the same similarity functions as the context model. The distance between item i and Prototype A is given by

$$d_{iA} = \sum w_m \cdot |x_{im} - P_{Am}|, \tag{6}$$

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

$$s_{iA} = \exp(-c \cdot d_{iA}), \tag{7}$$

where c is the sensitivity parameter. Analogous equations are used to compute the similarity of item i to Prototype B. The probability with which item i is classified into Category A is given by

$$P(A|i) = \frac{g}{2} + (1-g) \cdot \frac{s_{iA}}{s_{iA} + s_{iB}}. \tag{8}$$

A response-scaling parameter γ could be added to the multiplicative-similarity prototype model in a manner analogous to the context model (Equation 1). Critically, however, within the framework of the multiplicative-similarity prototype model, the γ parameter cannot be estimated separately from the sensitivity parameter c. That is, without loss of generality, the γ parameter in the multiplicative-prototype model can be set equal to one and its influence absorbed by the sensitivity parameter c. To see why, let

$$P(A|i) = \frac{{s_{iA}}^{\gamma}}{{s_{iA}}^{\gamma} + {s_{iB}}^{\gamma}}. \tag{9}$$

Note that ${s_{iA}}^{\gamma} = [\exp(-c \cdot d_{iA})]^{\gamma} = \exp(-c \cdot \gamma \cdot d_{iA}) = \exp(-c' \cdot d_{iA})$, where $c' = c \cdot \gamma$. Thus, the role of the γ response-scaling parameter is already implicit in the multiplicative-prototype model.[1] We discuss this crucial issue in greater depth in the section The Response-Scaling Parameter γ.

Mixed Model

As discussed by Smith and Minda (1998, 2000), with extended training on a fixed set of exemplars, there is evidence that stored exemplars come to play an important role in influencing classification performance. However, these researchers hypothesized that the memories of these exemplars are "all-or-none" in the sense that they are used only for purposes of classifying the training instances themselves. Generalizations to new items are assumed to be based solely on similarity comparisons to the prototypes. Smith and Minda (2000; pp. 9–10, 13–18) argued strongly that this all-or-none exemplar memory assumption is simpler and psychologically more plausible than is the exemplar-similarity process formalized in the context model. Thus, according to Smith et al.'s (1997; Minda & Smith, 2001; Smith & Minda, 1998, 2000) proposed mixed model of classification, the probability that an old training instance from Category A is correctly classified into Category A is given by

$$P(A|i) = e + (1-e) \cdot \frac{g}{2} + (1-e) \cdot (1-g) \cdot \text{proto}(A|i), \tag{10}$$

where e ($0 \leq e \leq 1$) is the probability that the observer has formed a memory trace for an individual exemplar, g is the guessing probability, and proto (A | i) denotes the probability that item i is classified into Category A by the prototype process. If item i is a new item that was not presented during training, then $e = 0$ and classification is based solely on a combination of the guessing and prototype processes. In our research we considered two different versions of the mixture model, one that assumed an additive-prototype process and the other that assumed a multiplicative-prototype process.

The Response-Scaling Parameter γ

In the original formulation of the exemplar-based context model, Medin and Schaffer (1978) assumed that classification responses were based on matching to the relative summed similarities of each category. That is, they assumed $\gamma = 1$ in the response rule formalized in Equation 1. There was never any strong theoretical justification for the use of this response rule, however.[2] For example, Medin and Smith (1981) wrote, "The best defense of the response rule is that it is a fair approximation and that it seems to work" (p. 250). Indeed, when one uses the context model to fit classification data averaged across subjects, the model

[1] By contrast, because the exemplar model involves a summing of similarities (see Equation 1), the γ response-scaling parameter is mathematically distinct from the sensitivity parameter in that model. We note, however, that even in the exemplar model, there are situations in which it is difficult to estimate separately the values of these two parameters.

[2] One possible source of justification is the relation between the context-model response rule and the classic similarity-choice model (SCM) of stimulus identification (Luce, 1963; Shepard, 1957), which does not include the γ response-scaling parameter. According to the SCM, the probability that stimulus i is identified as stimulus j is given by $P(j \mid i) = \frac{b_j \eta^*_{ij}}{\sum b_k \eta^*_{ik}}$, where b_j is the bias for response j and η^*_{ij} is the similarity between stimuli i and j. For the same reason as in the multiplicative-prototype model, however, the γ response-scaling parameter is nonidentifiable within the framework of the SCM. Thus, the model $P(j \mid i) = \frac{b_j \eta^{\gamma}_{ij}}{\sum b_k \eta^{\gamma}_{ik}}$ is formally identical to the standard version with $\eta^*_{ij} = \eta^{\gamma}_{ij}$. If similarity in the SCM is assumed to be an exponential decreasing function of distance in psychological space (Shepard, 1957), then the γ response-scaling parameter cannot be estimated separately from the sensitivity parameter c.

with $\gamma = 1$ tends to fit quite well. However, numerous independent lines of evidence since then have indicated that when fitting individual subject data with the exemplar model, it is important to allow γ to take on values greater than 1 (e.g., Maddox & Ashby, 1993; McKinley & Nosofsky, 1995; Nosofsky & Palmeri, 1997; see also Nosofsky, 1991a, for a slightly different version of a deterministic exemplar model). As previously discussed in the Review of the Models section, the γ response-scaling parameter describes the extent to which observers use deterministic versus probabilistic response strategies, and values of $\gamma > 1$ allow the exemplar model to account for the levels of deterministic responding that are often evidenced by individual subjects (Ashby & Gott, 1988; Lovett, 1998; Ross & Murphy, 1996).

A process interpretation for the emergence of the γ parameter was provided by Nosofsky and Palmeri (1997) in the form of their exemplar-based random-walk (EBRW) model of classification. In that model, γ corresponds to the magnitude of the response criteria in a random-walk process in which exemplars are retrieved from memory; that is, it corresponds to the amount of exemplar-based evidence that needs to be obtained before an observer will initiate a response (see Nosofsky & Palmeri, 1997, p. 291).

Indeed, essentially all modern models of classification include a response-scaling parameter that is analogous to γ, including the response-scaling constant φ in Kruschke's (1992) attention learning covering map (ALCOVE) model (see Kruschke, 1992, p. 24), the goal-value parameter G and response-noise parameter t in Anderson and Lebiere's (1998) adaptive control of thought–revised (ACT–R) theory (see Lovett, 1998, pp. 256–257, 276–277), and the criterial-noise parameter σ_c^2 in Ashby and Maddox's (1993) decision-bound theory (see Ashby & Maddox, 1993, pp. 377–378).

Smith et al. have strongly criticized the role of the response-scaling parameter γ in the context model. For example, Smith and Minda (1998) argued that γ functions to allow the exemplar model to mimic what is in truth a prototype-based process, suggesting that the response-scaling parameter "can be a prototype in exemplar clothing" (p. 1431). Likewise, Minda and Smith (2001, p. 794) argued that including the γ parameter in the context model left the competing exemplar and prototype models unbalanced in their assumptions and parameters and endowed the exemplar model with too much flexibility. These arguments did not recognize, however, that the multiplicative-similarity prototype model itself already contains the formal flexibility that in essence accommodates a response scaling parameter (see our previous section, Review of the Models). Not including the γ parameter in the exemplar model creates a lack of balance in favor of the multiplicative-similarity prototype model, at least with respect to response-scaling processes.

Finally, although the additive-prototype model does not currently include a response-scaling parameter, one of our initial goals for the present article is to demonstrate the limited generality of this proposed model, even if it is extended by the all-or-none exemplar-memory process proposed by Smith et al. (Minda & Smith, 2001; Smith & Minda, 1998, 2000). In our next section, we suggest that the additive-prototype model has fared well in Smith et al.'s experimental paradigm for a very special reason, and we argue that somewhat more challenging paradigms can quickly reveal the inadequacy of this model.

The Prototype-Plus-Exception Structure

The main experimental paradigm used by Smith et al. for demonstrating a purported qualitative advantage of the prototype models over the exemplar model is shown in Table 1. There are seven training exemplars in each of two categories, A and B. Each exemplar varies along six binary-valued dimensions, with logical value 1 on each dimension tending to indicate Category A and logical value 2 tending to indicate Category B. In this design, Stimulus 1 (*111111*) is the prototype of Category A, whereas Stimulus 8 (*222222*) is the prototype of Category B. Note also that each category contains an "exception" item. Stimulus 7 (*222212*) is the exception stimulus that belongs to Category A, whereas Stimulus 14 (*111211*) is the exception stimulus that belongs to Category B. All of the remaining exemplars in each category differ from their respective prototypes along just a single dimension. We refer to these items as the *low distortions* of the prototype.

Smith et al. (1997, Smith & Minda, 1998) tested designs in which subjects learned to classify these 14 training instances into each of the two categories. They then fitted the various competing models to the classification learning data of each of their individual subjects. They found that at early stages of learning, the prototype models provided better fits to the classification data than did the exemplar model (with $\gamma = 1$), at least for certain subsets of the subjects who were tested.

To characterize the limitations of the exemplar model, Smith et al. (1997) created composite plots showing the observed and predicted percentage of correct responses for each individual stimulus. (Although the models were fitted to the data of each individual subject, the composite plots were created by averaging across the observed and predicted data of the subjects.) An example of one such plot is provided in Figure 1, which shows the predictions from the exemplar model and the observed data from a subset of subjects who showed a "prototype-based" pattern of performance. The key aspect of the plot emphasized by Smith et al. (1997) is that the exemplar model underestimates the predicted percentage of correct responses for the prototype patterns (Stimuli 1 and 8) and overestimates the predicted percentage of correct responses for the exception patterns (Stimuli 7 and 14). By contrast, these systematic deviations were not present in the prototype-model fits to this subset of subjects.

Consider the following highly simplified process that might be taking place when observers learn to classify the stimuli in this paradigm. Numerous studies suggest that at early stages of learning involving separable-dimension stimuli, observers may engage

Table 1
The Prototype-Plus-Exception Structure Tested by Smith et al. (1997; Smith & Minda, 1998) and in Experiments 2 and 3 of the Present Research

Category A	Category B
1. 111111	8. 222222
2. 211111	9. 122222
3. 121111	10. 212222
4. 112111	11. 221222
5. 111121	12. 222122
6. 111112	13. 222221
7. 222212	14. 111211

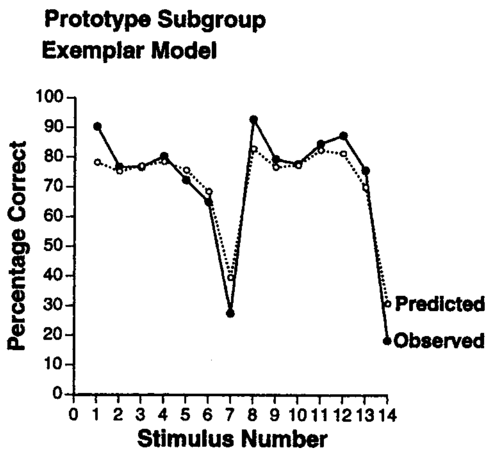

Figure 1. The solid line shows the average observed percentage of correct category decisions displayed by the "prototype subjects" for each of the 14 stimuli in Smith et al.'s (1997) Experiment 1 (nonlinearly separable, easy condition). The dotted line shows the average percentage of correct category decisions for each stimulus predicted by the exemplar-based context model with $\gamma = 1$. From "Straight Talk About Linear Separability," by J. D. Smith, M. J. Murray, and J. P. Minda, 1997, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* p. 669. Copyright 1997 by the American Psychological Association. Adapted with permission of the authors.

in hypothesis-testing behavior in which they sample rules along individual dimensions (Levine, 1975; Medin & Schaffer, 1978; Nosofsky, Palmeri, & McKinley, 1994; Trabasso & Bower, 1968). Thus, suppose that an observer formed a rule along Dimension 1. Any stimulus with value 1 along Dimension 1 would be classified into Category A, whereas any stimulus with value 2 along Dimension 1 would be classified into Category B. Each individual subject, however, might form a rule along a different dimension. For simplicity, we assume that the rule dimension is chosen randomly for each subject. Finally, analogous to the guessing process formalized in the prototype models, subjects might guess randomly between the categories on some proportion g of the trials. The composite predictions generated by this highly simplified model with $g = .10$ are displayed in Figure 2. Remarkably, this highly simplified "rule" model, in which each observer is making use of information along just a single dimension, yields the type of composite profile reported by Smith et al. (1997) in their studies. (A more precise quantitative fit could be achieved by estimating as free parameters the probabilities with which individual subjects adopted rules along the different dimensions.) Later in our article, we analyze the data from individual observers to provide evidence that a process not very different from this highly simplified rule-based process may indeed underlie classification behavior at the very early stages of learning.

This pattern of classification responding, however, does not constitute evidence that people are using prototype-based strategies rather than exemplar-based strategies, as Smith et al. claim. The additive-prototype and multiplicative-prototype models would fit such individual performances simply by placing all of their attention weight on the single rule dimension. Likewise, the exemplar model would fit such performances equally well by placing all of its attention weight on this single dimension. (Single-dimension rules of this form are a very special case of the types of behavior that the selective-attention exemplar model was designed to explain—see, e.g., Medin & Schaffer, 1978; Nosofsky, 1984, 1986, 1991b.) However, the value of γ in the exemplar model would need to be set at a sufficiently high value to account for the deterministic pattern of responding assumed in the rule-based process.

As discussed previously, in all of their previous quantitative tests of the exemplar model, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) held γ fixed at 1. Suppose that an observer attended selectively to Dimension 1 and gave zero attention to the remaining dimensions. Note from Table 1 that five of the seven exemplars from Category A have logical value 1 on Dimension 1, whereas two of the seven exemplars from Category B have logical value 1 on this dimension. Applying the exemplar model with $\gamma = 1$, one can directly show that the model predicts, for example, that the prototype of Category A will be classified correctly with maximum probability $5/(5 + 2) = .71$, which is far less than the observed average proportion correct of over .90 reported in the composite plot of Figure 1. This underprediction, we believe, lies at the heart of essentially all of Smith et al.'s demonstrations of the limitations of the exemplar model.

Our view, however, is that constraining the exemplar model to match to the relative summed similarities, as is the case in the $\gamma = 1$ model, is arbitrary. From a psychological point of view, an observer at early stages of learning may remember that five exemplars from Category A had value 1 on Dimension 1 and that two exemplars from Category B had value 1 on Dimension 1. When tested with an object that has value 1 on Dimension 1, a subject's reasonable response strategy would simply be to classify it into Category A, as would be predicted by a deterministic version of the exemplar model with γ set at a sufficiently high level.

The plan in the remainder of our article is as follows. We start by reporting something of a "demonstration" experiment to show clearly that the additive-prototype model is inadequate as a model of classification performance, even if augmented with an all-or-

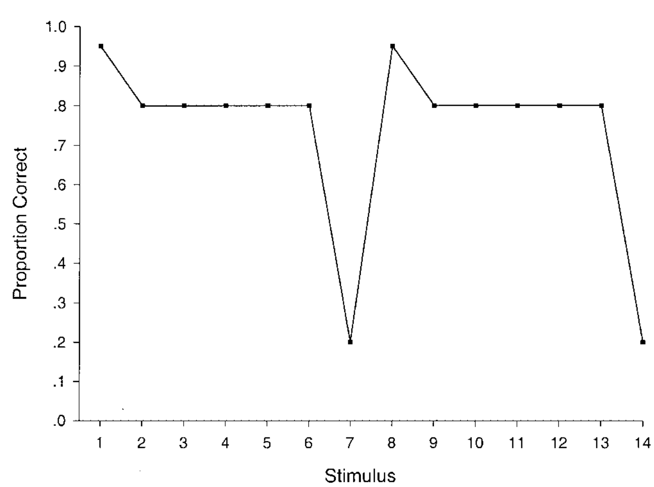

Figure 2. Composite predictions generated by a highly simplified single-dimension rule model for the prototype-plus-exception structure tested by Smith et al. (1997; Smith & Minda, 1998). Compare the form of the plot to the observed data in Figure 1.

none exemplar memory process. Just as is the case with the exemplar model and the multiplicative-prototype model, the additive-prototype model too needs to be endowed with a response-scaling mechanism.

Next, we conduct a replication and extension of the key experimental paradigm used by Smith et al. that purportedly found evidence for a prototype-abstraction process. Detailed analyses of the individual subject data obtained at early stages of learning suggest that rather than having adopted a broad-based prototype, as might be conveyed by inspecting the composite plot in Figure 1, individual subjects are indeed attending selectively to a small subset of the stimulus dimensions, very much in the spirit of the simple rule-based process used to generate our Figure 2. The predictions from exemplar and prototype models can simply not be distinguished in such a situation.

In a final experiment, we follow Smith and Minda (1998) by having subjects learn the prototype-plus-exception structure in a more extended training session. To learn the category structure, subjects must spread their attention to other relevant dimensions of the objects. Critically, however, whereas Smith and Minda (1998) tested observers on only the old training instances, in our experiment we extend the paradigm by also testing how observers generalize to new transfer stimuli. This extension allows the predictions from the exemplar and prototype models to be sharply distinguished. Model-based analyses of the individual subject data from such a paradigm reveal that the standard exemplar-generalization model provides a dramatically better account of the observed classification performance than do the alternative prototype models, even ones that are augmented with an all-or-none exemplar-memory process.

Experiment 1

The purpose of Experiment 1 was to provide a simple demonstration of the limitations of the additive-prototype model, even if this model is augmented with the all-or-none exemplar-memory process.

The key manipulation in our experiment was to train individual observers on only the actual prototypes of each category. The stimuli were composed of six binary-valued dimensions. To lessen the possibility that observers would attend selectively to only a single dimension (or to just a few dimensions), we provided observers with explicit instructions that each dimension of the objects was equally important. Finally, in a transfer phase, we tested subjects not only with the prototypes on which they were trained but also with various new transfer stimuli that were distortions of the prototypes. Low distortions differed by one dimension value from their prototype and medium distortions differed by two dimension values. To ensure highly motivated subjects, we provided payoffs for good performance. The subjects were not provided with feedback on the various transfer patterns but were told that there were right and wrong answers and that they would be paid on the basis of their performance.

Because only the prototypes were presented as training exemplars, the context model and the multiplicative-prototype model are formally identical in this experimental paradigm. The purpose of the paradigm was to sharply distinguish the multiplicative-similarity models from the additive-prototype model. The intuition here is straightforward. Presumably, an observer wishing to maximize payoffs would classify an object into Category A if it were more similar to the Category A prototype than to the Category B prototype, and vice versa for Category B objects. Thus, observers should respond in near deterministic fashion, with correct choice probabilities close to 1.0 for all patterns (as long as the observers distribute their attention across all of the dimensions of the objects). The multiplicative-similarity models account for such performance in a straightforward fashion simply by setting the response-scaling (or sensitivity) parameter at a sufficiently high level. By contrast, without a response-scaling process, the additive-prototype model is constrained to make much different predictions. Assuming, for example, that observers distribute attention equally among the six dimensions (i.e., so that each w_m attention-weight parameter in Equation 4 has a value of .167), then low distortions are predicted to be correctly classified with maximum probability .833 and medium distortions are predicted to be correctly classified with maximum probability .667. Furthermore, making allowance for the idea that observers supplement the prototype process by forming all-or-none memories for the training patterns does nothing to save the additive-prototype model because the only training patterns are the prototypes themselves.

Method

Subjects. Eight Indiana University graduate students were paid $7 for participating in the experiment. In addition, subjects were given a bonus of $3 if they achieved at least 95% correct in the task. We tested graduate students and used payoffs because achieving a clear demonstration of the inadequacies of the additive-prototype model requires that subjects perform at high levels. None of the subjects was aware of the issues under investigation in this study.

Stimuli. The stimuli were the six-letter nonsense words used by Smith et al. (1997; Smith & Minda, 1998) in some of their experiments. Each letter in the nonsense strings is presumed to constitute a feature. There were four prototype pairs (*HAFUDO–NIVETY*; *GAFUZI–WYSERO*; *BANULI–KEPIRO*; and *LOTINA–GERUPY*). Low distortions of each prototype were created by substituting one feature that was prototypical of one category with the feature that was prototypical of the other category. Medium distortions were created by substituting two features. There was a total of 44 different stimuli used in the experiment (1 prototype, 6 low distortions, and 15 medium distortions from each of the two categories).

Procedure. Subjects were randomly assigned to one of the four prototype pairs. At the outset of the experiment, the subjects studied the two category prototypes and were told that all of the words in the experiment were derived from these two nonsense words. Subjects were then tested on the 44 members of the nonsense-word categories. Each item appeared on the screen and the subjects were instructed to press the key that corresponded to the correct category. Subjects received feedback on only those trials involving presentations of the prototypes. There was a total of four test blocks in the experiment. In each block, the prototype was shown four times and each of the remaining stimuli was shown once. The order of presentation of the stimuli was randomized within each block for each subject.

Results and Theoretical Analysis

The observed proportions of correct responses for the prototypes, low distortions, and medium distortions are reported separately for each of the eight observers in Table 2. (The data here are averaged across the individual tokens of these main item types.) Inspection of the table reveals immediately that the observed correct classification proportions for all eight observers are far

Table 2
Classification Probabilities and Summary Fits in Experiment 1

Subject	Model	Pattern type: Proto	Low	Medium	−ln L
1	Obs.	1.0	.94	.78	
	Exemp.	1.0	.95	.77	22.9
	Add-prot.	1.0	.83	.67	33.7
2	Obs.	1.0	1.00	.95	
	Exemp.	1.0	1.00	.95	10.4
	Add-prot.	1.0	.83	.67	52.6
3	Obs.	1.0	.98	.96	
	Exemp.	1.0	1.00	.94	16.8
	Add-prot.	1.0	.83	.67	53.5
4	Obs.	1.0	1.00	.98	
	Exemp.	1.0	1.00	.97	7.1
	Add-prot.	1.0	.83	.67	55.0
5	Obs.	1.0	1.00	.98	
	Exemp.	1.0	1.00	.98	3.4
	Add-prot.	1.0	.83	.67	54.6
6	Obs.	1.0	1.00	.98	
	Exemp.	1.0	1.00	.98	3.3
	Add-prot.	1.0	.83	.67	55.7
7	Obs.	1.0	1.00	.84	
	Exemp.	1.0	.99	.86	18.2
	Add-prot.	1.0	.83	.67	41.0
8	Obs.	1.0	1.00	.93	
	Exemp.	1.0	1.00	.93	14.9
	Add-prot.	1.0	.83	.67	53.8

Note. Smaller values of −ln L indicate a better fit of a model. Proto = Prototypes; Low = Low distortions; Medium = Medium distortions; −ln L = negative log-likelihood; Obs. = observed correct proportions; Exemp. = predictions from exemplar model; Add-prot. = predictions from additive-prototype model.

higher than can be predicted by the additive-prototype model. Indeed, five of the eight observers correctly classified even the medium distortions with probability greater than or equal to .95.

We fitted the exemplar model and the additive-prototype model to each individual subject's classification data by using a maximum-likelihood criterion. The criterion was to maximize the log-likelihood function

$$\ln L = \sum \ln N_i! - \sum\sum \ln f_{ij}! + \sum\sum f_{ij} \ln p_{ij},$$

where N_i denotes the frequency with which stimulus i was presented, f_{ij} denotes the frequency with which the subject classified stimulus i into Category j, and p_{ij} denotes the predicted probability with which the subject classified stimulus i into Category j. This likelihood function is computed under the assumption that the responses for each stimulus are binomially distributed into the two categories and that the distributions for each stimulus are independent.

The predictions of the models, averaged across the individual tokens of the main item types, are reported along with the observed data in Table 2. We also report the summary fits of the models to each individual subject's data. The exemplar model's (and multiplicative-prototype model's) predictions pinpoint the observed data, whereas the predictions from the additive-prototype model fall completely short.

Discussion

These results provide a simple demonstration that the additive-prototype model advanced by Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998), even the mixed version of the model that allows for all-or-none exemplar memories, is inadequate as a general model of classification performance. Although Smith et al. have strongly criticized the context model's use of the γ response-scaling parameter, the present results demonstrate that the additive-prototype model is in just as much need of an analogous response-scaling process.

One approach to extending the additive-prototype model with a response-scaling mechanism is to exponentiate the Category A and Category B evidence terms and enter them into a response-ratio rule (cf. Kruschke, 1992):

$$P(A|i) = \frac{\exp(\varphi E_{i,A})}{\exp(\varphi E_{i,A}) + \exp(\varphi E_{i,B})}, \quad (11)$$

where φ is a response-scaling parameter. In the Appendix we show that this extension yields a model that is formally identical to the multiplicative-prototype model. Thus, our ensuing tests of the multiplicative-prototype model may be alternately viewed as tests of a version of the additive-prototype model extended with a response-scaling mechanism.[3]

Given the obvious inadequacies of the version of the additive-prototype model without a response-scaling mechanism, we do not consider it further in the remainder of this article. Although we fitted it to the data in our subsequent experiments, it never fitted better, and often fitted substantially worse, than did the multiplicative-prototype model, even when these models were extended with the all-or-none exemplar-memory process. Our remaining experiments therefore focus on comparisons between the multiplicative-similarity exemplar and prototype models, both of which incorporate analogous response-scaling mechanisms in their machinery.

Experiment 2

In the introduction to our article, we raised the possibility that at early stages of learning in Smith et al.'s (1997; Smith & Minda, 1998) experimental paradigms, observers may be attending to a small subset of the dimensions that compose the objects. As explained earlier, if observers adopt this type of selective attention strategy and also use deterministic response rules for making classification responses, then the special-case version of the exemplar model with $\gamma = 1$ will fail to fit the data. The purpose of Experiment 2 was to test this possibility by replicating and extending Smith et al.'s (1997; Smith & Minda, 1998) original paradigm and conducting detailed modeling analyses of the classification performance of individual observers.

[3] In all of our ensuing tests, we also fitted the following extended version of the additive-prototype model:

$$P(A|i) = \frac{E^{\gamma}_{i,A}}{E^{\gamma}_{i,A} + E^{\gamma}_{i,B}},$$

where γ is a response-scaling parameter. In all cases, this version produced essentially the same fits as did the multiplicative-prototype model.

In their articles, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) did give some consideration to the possibility that subjects were using single-dimension rules at the early stages of classification learning. They rejected this possibility, however, on the grounds that certain rule-based models provided worse overall fits to their classification data than did the prototype models.

There are some important limitations of Smith et al.'s (1997; Minda & Smith, 2001; Smith & Minda, 1998) methods and analyses that need to be considered, however. First, in their studies Smith et al. fitted data that were obtained during the course of a learning sequence in which corrective feedback was continually provided. Even if individual subjects were using single-dimension rules as a basis for classification, it is unlikely that subjects would maintain the same rule throughout the entire learning sequence because no single-dimension rule is available that allows for perfect performance. Thus, subjects would likely shift attention to alternative dimensions in their search for a single-dimension rule. The averaged data that are produced by cumulating across the trials of such a learning sequence will therefore not reflect the single-dimension strategies that may underlie performance. To remedy this difficulty, in the present experiment we trained subjects on four blocks of learning trials and then followed this training phase with a transfer phase. By withholding feedback during the transfer phase, we allowed for the possibility that individual observers would maintain whatever classification strategy had been developed to that point and would not continually shift attention to new dimensions of the objects. This technique is similar to the "blank trials" method used in classic research for investigating hypothesis-testing behavior (Levine, 1966).

A second limitation associated with Smith et al.'s analyses is that they showed only that the absolute fit of the rule model was worse than that of the prototype models. Because the rule model that they investigated is a highly constrained, special case of the prototype model in which all attention is given to a single dimension, it is impossible for the absolute fit of the rule model to be better than that of the prototype model. Therefore, little information is provided by reporting that the absolute fit value of the rule model was worse than that of the prototype model. In our ensuing analyses, rather than comparing the models on their absolute fit values, we test whether restricted models that assume attention to only a limited number of dimensions fit the classification data significantly worse than do full models that allow attention to all dimensions of the objects.

Finally, to gain more diagnostic information regarding subjects' classification performance, rather than presenting only the old training items at time of transfer, we also presented subjects with new low and medium distortions of the prototype, as well as with neutral patterns that were equally similar to the prototypes of each category. By requiring the alternative models to fit the choice probabilities of both the training items and the various new transfer patterns, we obtained more rigorous tests of the predictions from the models.

Method

Subjects. Forty Indiana University undergraduates participated to partially fulfill a requirement of an introductory psychology class. A bonus of $15 was paid to the 2 subjects who achieved the highest accuracy in the experiment.

Stimuli. The stimuli were drawings of bug-like creatures used previously by Smith and Minda (1998). The stimuli differed along six binary-valued features: head shape (round or oval), eye type (red open eye or green half-closed eye), antenna type (curved forward with purple dot or straight back with orange dot), body length (short or long), leg height (short or long), and feet type (blue semi-circle or gray triangle). The assignment of physical dimensions to the logical structure of the stimuli was randomized for each subject. In addition, following Smith and Minda (1998), four randomizations were used in assigning physical feature values to the logical values 1 or 2 along each dimension. For example, for different subjects, logical value 1 on the head-type dimension might result in either an oval head or a circular head. The stimuli are described in more detail in Smith and Minda (1998).

The set of training stimuli consisted of the 14 items whose abstract structure is listed in Table 1. The transfer set consisted of all 64 stimuli (including the old training stimuli) that can be constructed from the combination of 6 binary-valued dimensions. In particular, the transfer set included the prototype of each category, 6 low distortions of each prototype (items that differed by one feature from each prototype), 15 medium distortions of each prototype (items that differed by two features), and a total of 20 neutral items that differed by three features from each prototype.

Procedure. The subjects were told that their task would be to classify cartoon bugs into one of two categories. The various features of the bugs were listed and described prior to the start of the training phase. The subjects were instructed to look carefully at each bug and to classify it into one of the two categories and were told that the 2 most accurate subjects would receive a bonus of $15.

In the training phase, there were four blocks of the 14 training items. On each trial, a stimulus appeared on the screen with the text "Category 1 or 2?" underneath it to remind subjects of their task. Subjects received feedback after each response. The stimulus remained visible until the end of the feedback.

The training phase was immediately followed by a transfer phase in which subjects were presented with four blocks of the 64 stimuli. Order of presentation was randomized within each block. No feedback was provided. The subjects were instructed that they would be presented with both old and new stimuli and that their goal was to achieve as many correct classifications as possible. The instructions also explained that we were interested in discovering what strategy subjects had developed by the end of the learning phase, so they should try to use that same strategy consistently during transfer.

Results and Theoretical Analyses

The key question addressed in this experiment is the extent to which individual observers focused attention on a limited number of the dimensions composing the training exemplars. Our general plan for addressing this issue is to first fit the full version of the exemplar model to the individual subjects' classification data. Assuming that the full version of the model provides a good description of the data, we then proceed to fit various restricted versions of the exemplar model in which we place constraints on the number of attended dimensions. For example, in a single-dimension version of the model, we assume that all attention weight is placed on a single dimension, with zero attention weight placed on the remaining dimensions. We then determine whether the single-dimension model provides a significantly worse fit to each individual subject's data than does the full version of the model that allows attention to all six dimensions.

Before proceeding to these tests of the restricted attention models, we first verify that the full version of the exemplar model provides an adequate framework for analyzing the data. In Table 3,

Table 3
Maximum-Likelihood Fits (in Terms of −ln L) of the Models to the Classification Transfer Data Obtained in Experiment 2

Sub.	Full proto.	Full exemplar	3D-exemplar	2D-exemplar	1D-exemplar	Matching exemplar
1	22.787	22.825	22.866	23.240	31.112*	42.479
2	13.235	13.407	13.407	13.591	16.040	41.964
3	37.354	37.354	37.354	37.477	42.147	78.951
4	75.470	75.148	75.580	78.006	90.969*	77.344
5	10.288	10.548	13.820	89.022*	140.568*	83.243
6	60.347	60.797	61.285	61.632	62.272	60.863
7	17.595	17.595	17.595	17.595	17.698	83.783
8	32.500	28.695	28.695	29.547	32.783	43.049
9	56.614	58.393	100.382*	108.341*	120.271*	65.844
10	24.516	26.130	26.411	27.510	28.365	81.170
11	44.617	44.641	44.983	47.976	53.480*	74.300
12	53.302	53.322	54.863	55.627	59.107*	69.630
13	16.406	16.447	16.658	16.716	17.698	41.303
14	43.607	43.489	43.489	43.605	44.497	48.237
15	4.709	4.829	6.192	7.546	8.924	40.154
16	44.950	44.698	45.001	46.843	48.340	75.713
17	81.402	82.550	82.688	83.567	84.517	82.927
18	22.904	22.911	23.127	23.431	24.509	42.423
19	33.533	33.534	33.534	36.518	44.422*	76.736
20	32.048	32.453	32.453	32.618	32.783	45.282
21	41.236	40.541	40.763	41.158	42.147	47.555
22	62.439	62.447	63.012	63.427	64.611	74.726
23	75.583	73.557	73.557	73.602	77.025	73.837
24	74.514	74.736	74.736	78.805	84.443*	77.979
25	68.948	68.627	68.627	70.898	104.535*	70.995
26	15.262	16.810	17.421	18.644	20.132	83.166
27	7.529	6.127	6.127	6.127	8.924	40.053
28	31.709	32.039	33.318	34.236	37.295	44.976
29	54.792	39.150	39.469	46.162*	144.577*	46.682
30	51.627	51.588	52.832	56.729*	81.759*	72.274
31	53.422	54.697	65.868*	84.268*	106.939*	72.372
32	42.795	41.643	41.797	50.832*	66.528*	48.136
33	62.886	63.362	63.866	64.247	64.992	76.842
34	11.715	11.893	11.893	12.427	15.059	40.964
35	20.982	22.298	22.836	24.423	26.495	41.953
36	39.647	37.456	38.792	39.942	42.332	44.852
37	81.341	81.374	85.218	105.014*	127.082*	94.153
38	50.385	50.445	53.615	69.014*	91.341*	64.784
39	48.500	48.533	48.554	49.081	50.787	50.773
40	6.127	4.709	4.710	6.127	8.924	85.118
Mean	40.741	40.295	42.185	46.889	56.661	62.690

Note. Sub. = Subject no.; Full proto. = Full version of the multiplicative-prototype model; Full exemplar = Full version of the exemplar model; *n*D-exemplar = *n*-dimensional restricted versions of the exemplar model; Matching exemplar = restricted version of the exemplar model with $\gamma = 1$. Asterisks denote cases in which the *n*-dimensional restricted versions of the exemplar model fit significantly worse than does the full version.
* $p = .05$.

we report the log-likelihood fits of the full exemplar model to each of the 40 subjects' classification data (see column labeled "Full Exemplar"). As a source of comparison, the multiplicative-prototype model was fitted to these data as well. The resulting fits are also shown in Table 3 (see column labeled "Full Proto"). The table reveals that the exemplar and prototype models yield virtually identical fits to the data of the individual subjects. The differences between the fits of the models (mean $-\ln L = 40.3$ for the exemplar model, mean $-\ln L = 40.7$ for the prototype model) do not approach statistical significance, $t(39) = 1.05$.

To gain a sense of the absolute fit of the exemplar model, in Figure 3, we present a composite scatterplot of the observed Category A response proportions for each of the 64 patterns against the predicted Category A response proportions. Note that although the model was fitted to each individual subject's data, in this composite plot the observed and predicted proportions are computed by averaging across the results from the 40 observers. Inspection of the composite plot indicates that the exemplar model is achieving accurate quantitative predictions of the data. Another indication of the model's performance is that it accounted for an average of 87.0% of the response variance in the Category A response probabilities of the 40 individual observers. Likewise, the prototype model accounted for an average of 86.7% of the response variance.

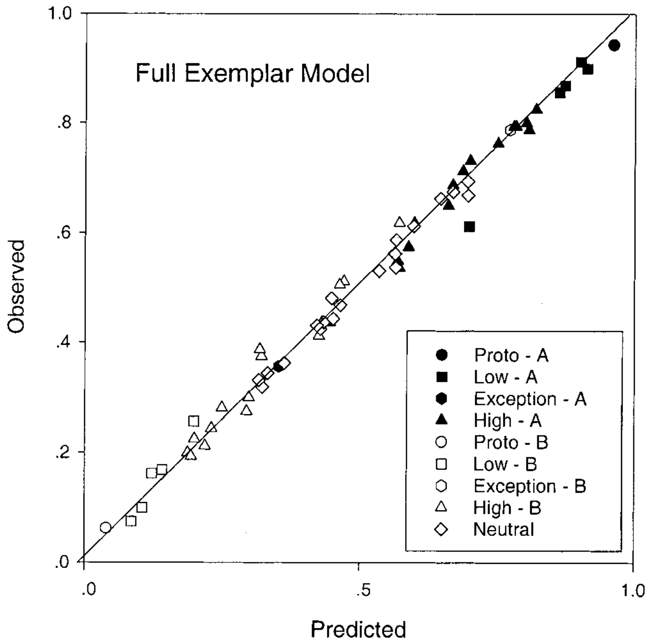

Figure 3. Composite scatterplot of the observed Category A response proportions for each of the 64 patterns against the predicted Category A response proportions from the full version of the exemplar-generalization model in Experiment 2. Proto = Prototype; Low = Low Distortion; Exception = Exception Pattern; High = High Distortion; Neutral = Neutral Pattern.

Having verified that the full version of the exemplar model provides a reasonable framework for analyzing the data, we now proceed to the tests of the restricted attention models. We fitted three different restricted versions of the model. In the first (1D-exemplar), we constrained the model such that all attention weight was given to a single dimension, with zero attention weight given to the remaining dimensions. In the second (2D-exemplar), we constrained the model such that all attention was split between two dimensions; and in the third (3D-exemplar), all attention was split among three dimensions. For simplicity in deriving the model fits, the dimensions that received attention were always those dimensions that had received the highest attention weights when the full version of the model was fitted to the data. These dimensions differed for each individual subject.

To test whether the restricted models provided significantly worse fits to each individual subject's data than did the full model, we used the method of likelihood-ratio testing (Wickens, 1982). Let $\ln L(F)$ denote the log-likelihood of the data when the full model is fitted to the data, and let $\ln L(R)$ denote the log-likelihood fit of the restricted model. Assuming that the restricted model is correct, then the statistic $G^2 = -2\ [\ln L(R) - \ln L(F)]$ has an approximate chi-square distribution with degrees of freedom equal to the number of parameters constrained in moving from the full to the restricted model.[4] If the observed value of G^2 exceeds the critical value, then one rejects the restricted model as fitting significantly worse than does the full model and concludes that at least some of the parameters were constrained inappropriately. We set alpha at .05 in conducting these tests.

The log-likelihood fits from the various restricted-attention exemplar models are reported along with those of the full exemplar model in Table 3. We denote with asterisks those cases in which the G^2 likelihood-ratio test leads one to reject the restricted model compared to the full model. For 25 of the 40 subjects, a single-dimension exemplar model, that is, a version of the exemplar model in which attention is given to only a single dimension, provides an adequate account of the data, that is, it is not rejected relative to the full version of the model. For 32 of the 40 subjects, the two-dimension version of the model is adequate, and for 38 of the 40 subjects, the three-dimension version is adequate.

To provide a better sense of these results, in Figure 4 we present the composite predictions from the three-dimension version of the model. Note that this model assumes that each subject attends to a maximum of three dimensions—for most of the subjects, almost all of the attention weight is placed on either one or two dimensions. Inspection of the composite scatterplot suggests that the low-dimension versions of the exemplar model are indeed providing an excellent account of the individual subjects' classification data.

Finally, in the last column of Table 3, we report the fits from the matching version of the exemplar model in which the response-scaling parameter is constrained at $\gamma = 1$, with all other parameters allowed to vary freely. Not surprisingly, as was found by Smith et al. (1997; Smith & Minda, 1998), the matching model provides relatively poor fits to the classification data of the individual subjects. Indeed, its fits are often even worse than those of the highly constrained single-dimension exemplar model. The reason for these poor fits is that the matching model is unable to capture the levels of deterministic responding exhibited by many of the individual subjects. To gain another sense of the importance of the γ response-scaling parameter, we examined the maximum-likelihood parameter estimates from the fits of the full version of the exemplar model. This analysis revealed that 39 of the 40 observers had γ estimates greater than 1, with the median estimate being $\gamma = 4.77$.

Discussion

In summary, the exemplar model provides an excellent account of the classification data observed at early stages of learning in the Smith et al. (1997; Smith & Minda, 1998) experimental paradigm, as long as the γ response-scaling parameter is allowed to vary. The fits of the exemplar model are essentially the same as those of the prototype model, which contains an implicit γ response-scaling parameter. Nosofsky and Johansen (2000) reported similar results when they re-analyzed some of the individual subject data from Smith et al.'s (1997) study.

In addition, the results of the present analyses provided evidence consistent with the idea that numerous individual observers were

[4] A technical issue regarding the number of constrained parameters in the restricted-attention models is that we do not specify a priori which dimensions receive zero weight. Nevertheless, we assume that the difference in the number of freely varying attention weights used in the two models being compared provides a good approximation to the degrees of freedom in the chi-square test. Additionally, in all of our model-fitting procedures, we set upper limits on the sensitivity and response-scaling parameters of $c = 100$ and $\gamma = 100$.

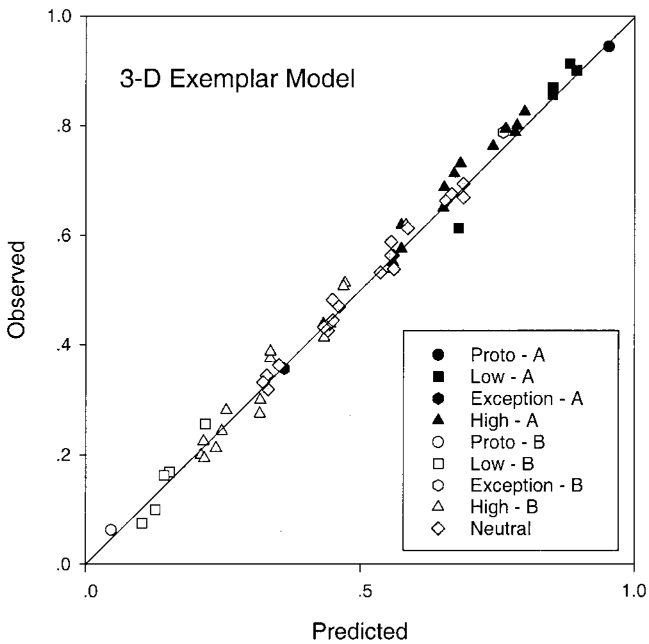

Figure 4. Composite scatterplot of the observed Category A response proportions for each of the 64 patterns against the predicted Category A response proportions from the 3-dimensional restricted version of the exemplar-generalization model in Experiment 2. Proto = Prototype; Low = Low Distortion; Exception = Exception Pattern; High = High Distortion; Neutral = Neutral Pattern.

attending selectively to a small subset of the dimensions composing the objects. At early stages of learning, subjects are not forming broad-based prototypes in this paradigm. The low-dimension rules that are being formed are equally well captured by versions of the exemplar and prototype models that assume selective attention to just a few dimensions and that make allowance for the operation of deterministic response strategies. Contrary to the claims of Smith et al., the behaviors of subjects during the early-learning stages of this paradigm do not discriminate between the predictions of exemplar and prototype models.

Experiment 3

The central goal of Experiment 3 was to develop strong contrasts between the predictions from the exemplar-generalization model and the mixed-prototype model advanced by Smith et al. (1997; Smith & Minda, 1998, 2000). Whereas in Experiment 2 we tested transfer performance after only four blocks of training trials, in Experiment 3 we conducted the transfer tests following 16 blocks of training. Smith et al. (1997; Smith & Minda, 1998, 2000) have acknowledged that following extensive training on a fixed set of items, memories for individual exemplars do play an important role in classification. However, they argued that the prototype representation continues to serve as the fundamental organizing principle of the category representation. The memories for the individual exemplars that are created are presumed to be all-or-none in the sense that they do not support any generalizations to new items. Such generalizations are presumed to be based on similarity to the prototypes.

Although Smith et al. (1997; Smith & Minda, 1998) demonstrated that this mixed-prototype model provided good accounts of classification performance at late stages of training, a limitation of their studies was that new transfer items were not included in the experimental designs. Rather, these researchers tested the ability of the mixed model to predict classification performance on only the old training items themselves. However, a major distinction between the exemplar-generalization model and the mixed-prototype model rests on these models' predictions of generalization to new transfer items. The exemplar model predicts that generalization is based on similarities to the stored exemplars, whereas the mixed-prototype model predicts that generalization is based on similarity to the prototype.

The prototype-plus-exception structure (see Table 1) designed by Smith et al. (1997; Smith & Minda, 1998) seems ideally suited for contrasting these models, as long as new items are included in the transfer tests. Specifically, consider new items that are similar to the old exception patterns that are stored in memory (i.e., Patterns 7 and 14 in the Table 1 structure). Because its exemplar-memory process is all-or-none, the mixed-prototype model tends to predict that such new items will be classified into the opposite category to which each exception belongs. That is, such new items will be classified in accord with the prototype structure of each category. By contrast, the exemplar-generalization model tends to predict that such new transfer items will be classified into the same category to which the respective exceptions belong. Although the precise predictions from the models depend on their parameter settings, the complete set of transfer stimuli provides multiple constraints such that the models cannot stray very far from the qualitative contrast outlined above.

Method

Subjects. Forty-three Indiana University undergraduates participated to partially fulfill a requirement of an introductory psychology class. Once again, we paid a bonus of $15 to the 2 subjects who achieved the highest accuracy in the experiment. None of the subjects had participated in the previous experiments.

Stimuli. The stimuli were the same as those used in Experiment 2.

Procedure. The procedure was identical to the one used in Experiment 2 except that in the current experiment, subjects trained on a total of 16 blocks of the training stimuli (instead of the four blocks in the previous experiment).

Results and Theoretical Analysis

Model-based analyses of the learners' data. Because the fundamental contrast between the exemplar-generalization model and the mixed-prototype model pertains only to those observers who learned the training structure, we focus our main analyses on those observers who achieved over 86% correct on the training items during the transfer phase. (Even if an observer responds perfectly to the prototypes and low distortions, this criterion can be achieved only if he or she makes at least one correct response for an exception item.) Twenty-two of the 43 subjects met this learning criterion. We report the modeling analyses for the nonlearners in a later section.

We fitted the exemplar-generalization model and the mixed-prototype model to the classification transfer data obtained for each of the 22 learners by using a maximum-likelihood criterion.

The log-likelihood fits for each of the models are reported in Table 4. As an auxiliary measure, we also report the percentage of variance in the Category A response probabilities that is accounted for by each model. The exemplar-generalization model provides a better log-likelihood fit than does the mixed-prototype model for 18 of the 22 learners. In a large number of cases, the improved fit provided by the exemplar-generalization model compared with the mixed-prototype model is quite dramatic. In no case does the mixed-prototype model provide a substantially better fit than does the exemplar-generalization model. Across the 22 learners, the log-likelihood value achieved by the exemplar-generalization model (mean $-\ln$ L $=$ 33.9) is significantly better than that achieved by the mixed-prototype model (mean $-\ln$ L $=$ 51.9), $t(21) = 3.89$, $p < .001$. The percent variance measure shows the same pattern of results.

It is instructive to consider the patterns of classification transfer data to gain an understanding of the reason for the better performance of the exemplar-generalization model compared to the mixed-prototype model. In the following analysis, we define the "neighbors" of exception-training Exemplars 7 and 14 as those novel transfer stimuli that differ from the exceptions along only a single dimension value. In general, the exemplar model tends to predict that the neighbors will be classified into the same category as the exception to which they are similar, whereas the mixed-prototype model tends to predict the opposite. The results for each of the neighbors, averaged across the observed and predicted classification probabilities from all 22 learners, are displayed in Figure 5. As can be seen, the exemplar model predicts well the classification probabilities for the neighbors. By contrast, the pattern of results is qualitatively inconsistent with the predictions from the mixed-prototype model. For both the Exception 7 neighbors and the Exception 14 neighbors, in four of five cases the mixed-prototype model predicts that these transfer stimuli will be classified into the opposite category from what was actually observed. (The reason that the mixed-prototype model predicts correctly the results for Transfer Stimuli *222112* and *111221* is that many subjects gave a good deal of attention weight to Dimensions 4 and 5. Transfer Stimuli *222112* and *111221* happen to match the corresponding prototypes on these particular dimensions.)

We emphasize that the summary results illustrated in Figure 5 are based on averaged data. Providing a full explanation of the pattern of results requires a specification of how individual observers distributed attention across the dimensions of the stimuli. We conducted additional analyses in which the neighbors were defined for each individual observer based on the particular dimensions that he or she attended. In these individual observer analyses, the superiority of the exemplar model over the mixed-prototype model was often even more pronounced than is illustrated by the composite plot in Figure 5.

Recall that Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) compared the mixed-prototype and exemplar-generalization models on their ability to fit the classification proportions associated with the training stimuli only. It is instructive to point out that had we followed this procedure, it would have yielded far less dramatic differences in quantitative fit between the models. When we fitted the models to the data from only the training instances, in 10 of 22 cases the mixed-prototype model gave slightly better log-likelihood fits than did the exemplar model, although the mean $-\ln$ L for the exemplar model (2.66) was still significantly smaller than that of the mixed-prototype model (3.51), $t(21) = 2.82$, $p < .05$. Both models yield extremely good fits for an uninteresting reason: As training proceeds, observers learn to classify all of the old instances with essentially perfect accuracy. The mixed-prototype model fits such performances simply by setting its exemplar-memory parameter e at a value near 1. Thus, testing the ability of the mixed-prototype model to generalize to new transfer stimuli is crucial for evaluating the utility of this model.

Model-based analyses of the nonlearners' data. For completeness, we also fitted the exemplar-generalization and mixed-prototype models to the data of the 21 observers who failed to meet the learning criterion. The model fits are reported in Table 5. With the exception of a few observers, there was little difference in the quantitative fits of the two models. The exemplar-generalization model provided a better fit for 11 of the 21 nonlearners, and its average log-likelihood fit ($-\ln$ L $=$ 47.1) was slightly better than that of the mixed-prototype model ($-\ln$ L $=$ 51.4). However, the differences in the log-likelihood fit values were not statistically significant, $t(20) = 1.52$, $p > .10$. In general, the behavior of most of the nonlearners was similar to those subjects from Experiment 2 who were trained on only four blocks of training trials. Because these subjects often focused attention on a single dimension, or on a small number of dimensions that were not perfectly diagnostic of category membership, they failed to learn to classify all of the

Table 4
Summary Fits of the Multiplicative-Prototype Model, the Mixed-Prototype Model, and the Exemplar-Generalization Model to the Learners' Transfer Data From Experiment 3

	Prototype		Mixed		Exemplar	
Sub.	$-\ln$ L	% Var	$-\ln$ L	% Var	$-\ln$ L	% Var
1	91.783	49.6	84.117	54.5	7.591	99.2
2	102.451	35.8	97.339	42.7	62.794	75.1
3	34.692	92.0	34.670	92.1	28.012	94.9
4	57.014	74.0	46.060	80.8	24.200	94.7
5	53.346	81.7	47.949	85.2	39.629	89.3
6	83.009	53.3	76.558	59.3	36.151	88.7
7	42.104	86.7	40.610	87.5	38.426	88.7
8	51.703	79.3	51.400	80.0	31.928	93.1
9	49.444	76.6	47.366	77.5	36.280	83.3
10	31.333	89.6	17.168	95.0	20.555	94.5
11	41.712	84.3	26.577	92.6	27.754	90.8
12	68.536	69.2	54.338	79.4	30.188	95.4
13	73.981	66.6	61.456	74.2	46.579	86.0
14	34.893	91.7	31.384	93.0	31.739	91.9
15	85.175	54.9	75.797	60.6	59.431	74.8
16	60.435	76.4	57.448	79.4	51.634	81.9
17	54.090	80.5	49.076	83.2	39.352	91.0
18	43.386	83.9	42.913	84.2	44.304	83.2
19	96.159	52.8	82.610	59.4	10.813	98.7
20	14.017	98.2	13.974	98.3	13.694	98.3
21	46.473	78.8	44.532	80.7	34.532	91.3
22	68.012	63.2	58.964	71.6	30.435	94.0
M	58.352	73.6	51.923	77.8	33.910	90.0

Note. Smaller values of $-\ln$ L reflect a better fit of a model. Sub. = Subject no.; $-\ln$ L = negative log-likelihood; % Var = percentage of variance accounted for; Prototype = Prototype model; Mixed = Mixed-prototype model; Exemplar = Exemplar-generalization model.

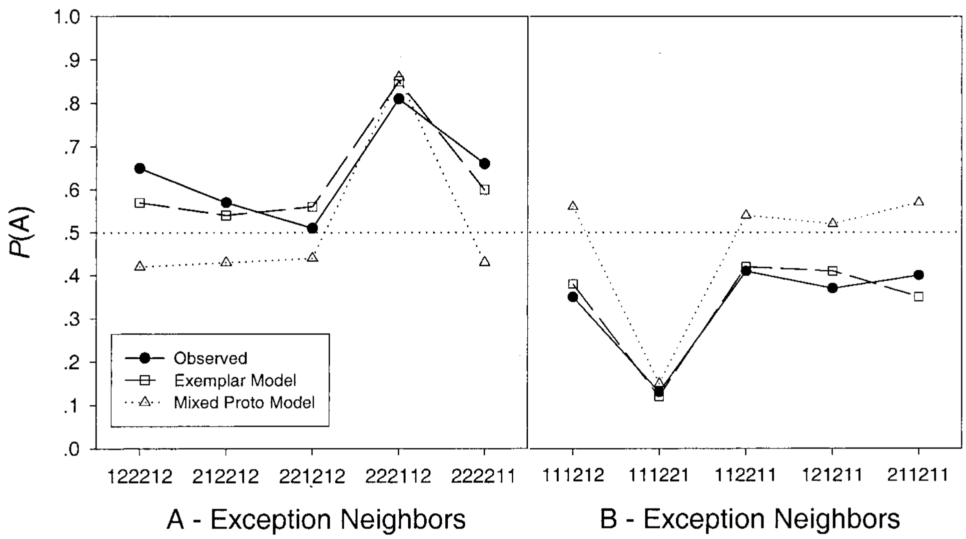

Figure 5. Averaged observed and predicted Category A response probabilities for the neighbors of the A exception and B exception in Experiment 3.

training exemplars into their appropriate categories. As a result, strong contrasts between the exemplar-generalization and mixed-prototype models could not be achieved for this subset of subjects.

Table 5
Summary Fits of the Multiplicative-Prototype Model, the Mixed-Prototype Model, and the Exemplar-Generalization Model to the Nonlearners' Transfer Data From Experiment 3

	Prototype		Mixed		Exemplar	
Sub.	−ln L	% Var	−ln L	% Var	−ln L	% Var
1	43.713	85.0	43.713	85.0	31.629	95.1
2	26.495	96.6	26.495	96.6	24.878	96.8
3	51.121	80.5	51.121	80.5	49.185	83.2
4	66.990	70.4	66.990	70.4	66.988	70.4
5	44.565	89.8	44.499	89.8	45.018	89.7
6	78.269	64.6	77.547	65.1	75.901	66.2
7	11.819	98.8	11.819	98.8	11.949	98.8
8	103.301	45.8	99.504	49.8	44.038	85.2
9	157.226	−0.2	157.226	−0.2	157.360	−0.3
10	86.200	52.6	84.431	54.1	85.810	52.9
11	54.396	72.8	54.396	72.8	32.991	82.3
12	17.390	98.1	17.390	98.1	16.518	98.1
13	17.142	98.0	17.142	98.0	18.319	97.9
14	3.064	99.7	3.064	99.7	2.354	99.7
15	76.707	48.2	76.684	48.3	76.682	48.0
16	65.795	73.4	65.795	73.4	66.565	72.8
17	7.530	99.3	7.530	99.3	7.815	99.3
18	8.924	99.2	8.924	99.2	8.924	99.2
19	31.079	95.0	31.080	95.0	31.171	94.9
20	70.772	58.5	70.772	58.5	70.951	58.4
21	63.755	73.9	62.261	74.0	63.898	73.6
M	51.726	76.2	51.352	76.5	47.092	79.2

Note. Smaller values of −ln L reflect a better fit of a model. Sub. = Subject no.; −ln L = negative log-likelihood; % Var = percentage of variance accounted for; Prototype = Prototype model; Mixed = Mixed-prototype model; Exemplar = Exemplar-generalization model.

Discussion

In summary, by including new stimuli in the transfer phase of the prototype-plus-exception structure, we were able to obtain strong contrasts between the predictions from the exemplar-generalization and mixed-prototype models at the later stages of learning. Generally speaking, the exemplar model predicted that items that were similar to the exceptions along attended dimensions would be classified into the exceptions' category, whereas the mixed-prototype model predicted that such items would be classified into the opposite category. The quantitative fit comparisons revealed a clear superiority for the predictions from the exemplar-generalization model, at least for those subjects who successfully learned the category structure by the completion of training. Inspection of the detailed predictions from the competing models suggested that the main reason for the superiority of the exemplar model was that it correctly predicted the patterns of generalization of new transfer stimuli to the exception items, a phenomenon that the mixed-prototype model was unable to explain.

General Discussion

Summary

In summary, in a recent series of studies, Smith and colleagues (Minda & Smith, 2001; Smith & Minda, 1998; Smith et al., 1997) presented evidence that they claimed challenges the predictions of the exemplar-based context model of classification and that supports the predictions of prototype models. The basis for these researchers' claims was that the prototype models provided better quantitative fits to certain sets of individual subject classification data than did the context model. However, in all of these quantitative-fit comparisons, Smith and colleagues considered the predictions from only a constrained version of the context model in which the response-scaling parameter γ was set at 1. This version of the context model assumes that observers make classification decisions by matching an item's relative summed similarity to the exemplars of the alternative categories. With regard to predicting individual observer data, the limitations of this matching version of the exemplar model are well known. As has been clearly demonstrated in past work, this version of the model fails to account for the more deterministic patterns of classification responding that are often evidenced at the individual observer level. The γ response-scaling parameter reflects the extent to which observers respond deterministically with the category that yields the largest exemplar-based similarity, rather than responding probabilistically by matching to the summed similarities.

Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) have criticized the role of the γ response-scaling parameter in the context model by arguing that it allows the context model to mimic what is in truth a prototype-based classification process. In our view, this criticism of the role of γ in the context model is unjustified. To our knowledge, all modern extant models of classification require a response-scaling process that is analogous to the one captured by the γ parameter. Indeed, as we demonstrated earlier, the multiplicative prototype model itself already includes the γ response-scaling parameter in implicit form. As explained in our Review of the Models section, in the multiplicative-prototype model, the γ response-scaling parameter cannot be estimated separately from the sensitivity parameter c. Thus, the multiplicative-prototype model advanced by Smith et al. is capable of producing any level of response determinism that is desired. Not allowing the context model an analogous response-scaling parameter creates a serious unbalance in favor of the prototype model. Thus, in our view, the comparisons conducted by Smith et al. confounded the issue of what is the nature of the underlying category representation with the issue of what type of response rule is used.

In its current form, the additive-prototype model advanced by Smith et al. (1997; Smith & Minda, 1998, 2000) does not include a response-scaling parameter, so this model may be viewed as being more parsimonious than the context model. However, the additive-prototype model is quite limited in generality. We hypothesized that it has worked well in some of the paradigms tested by Smith et al. for a special reason—namely, that at early stages of learning, numerous observers are using strategies that are akin to deterministic, single-dimension rules. The additive-prototype model can fit these deterministic patterns of responding by assigning all its attention weight to the single dimension. With slightly more challenging paradigms in which observers attend to multiple dimensions, we hypothesized that the inadequacies of the additive-prototype model would be easily revealed.

We tested these hypotheses in two main ways. First, we conducted a demonstration experiment in which observers were trained on only the prototypes of each category. To lessen the possibility that the observers would attend to just a small number of dimensions of the patterns, we provided explicit instructions that all dimensions of the objects were equally important, and motivated subjects by providing payoffs for good performance. Under these conditions, the additive-prototype model was constrained to predict that low and medium distortions of the prototype would be classified into their category with much lower probability than the prototypes themselves. In contrast to this prediction, observers correctly classified the low and medium distortions with extremely high probability, far greater than could be accounted for by the additive-prototype model. The experiment provided a simple demonstration that the additive-prototype model too needs to be augmented with some form of response-scaling process— contrary to Smith et al.'s arguments, this need is not unique to the exemplar-based context model.

Second, we tested explicitly the idea that under free-strategy conditions, numerous observers do indeed adopt single-dimension rules as a basis for classification at early stages of learning. In a direct replication and extension of the paradigm used by Smith et al. (1997; Smith & Minda, 1998), we found that at early stages of learning, 25 of the 40 observers adopted a classification strategy that was well described by models that assumed attention to only a single dimension. Indeed, the classification behavior of 38 of the 40 observers was well described by models that assumed attention to a maximum of three of the six stimulus dimensions. These performances were fit equally well by prototype and exemplar models that assumed attention to a subset of the stimulus dimensions. The experimental method of Smith et al. therefore appears to be nondiagnostic with respect to distinguishing these models.

Smith and Minda (1998, 2000) have acknowledged that following an extended learning sequence on a fixed set of exemplars, people may eventually form memories for those exemplars and use those memories to classify the old training instances. However, the organizing principle of the category structure is still assumed to be prototype based. Specifically, according to their mixed-prototype model of classification, there is some probability that stored exemplars are used to classify the old training instances, but the exemplar process is assumed to be all-or-none, with generalization to new items still assumed to be governed by similarity to the prototype.

However, Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998) did not provide rigorous tests to try to distinguish between the predictions from the mixed-prototype and exemplar-generalization models. They required the competing models to predict classification probabilities for only the old training instances themselves. A fundamental contrast between the models, however, concerns their predictions of how observers will generalize to new items. Therefore, to develop sharp contrasts between the models, we trained observers for an extended learning sequence on the prototype-plus-exception structure designed by Smith et al. (1997; Smith & Minda, 1998); however, at time of test, instead of presenting only the old training instances, we presented observers with a large set of new transfer patterns as well. In general, the exemplar-generalization model predicted that observers would classify patterns that were similar to the old exceptions into the same category to which the old exceptions belonged, whereas the mixed-prototype model predicted that observers would classify such patterns into the opposite category. In this experiment, the exemplar-generalization model provided a dramatically better quantitative fit to the individual subject data than did the mixed-prototype model. Furthermore, the main reason for the superiority in quantitative fit was that the exemplar model correctly predicted the patterns of generalization that were actually observed for the new transfer stimuli, whereas the mixed-prototype model did not. Stanton, Nosofsky, and Zaki (in press) reported a similar pattern of modeling results, except in a situation involving a linearly separable category structure instead of the prototype-plus-exception structure tested in the present work.

Issues of Model Flexibility

A key criticism that Smith et al. have raised regarding the role of the γ response-scaling parameter in the context model is that it makes the model too flexible, allowing it to fit data that are generated by alternative psychological processes. Such concerns about model flexibility are legitimate and are extremely important. Indeed, current work in mathematical psychology is focusing on the development of model-evaluation criteria that penalize formal models if they are too flexible (e.g., Myung, Forster, & Browne, 2000) and such work is likely to play a crucial role in future comparisons between prototype and exemplar models.

Nevertheless, these issues concerning measures of model flexibility are extremely complex and a good deal of caution is needed in applying and interpreting such measures. For example, it may well be that a highly constrained and simple model provides as good a quantitative fit as does a more complex model in some particular experimental paradigm. However, this simple model may be severely limited in generality, and it may fail dramatically in a slightly more complex paradigm. For example, the additive-prototype model fits well the data obtained at early stages of classification learning in Smith et al.'s (1997; Smith & Minda, 1998) experimental paradigms, but only because individual observers are apparently using simple single-dimension rules at these early learning stages. The same model fails dramatically in paradigms in which observers spread attention to multiple dimensions of the objects. It is important for models to not only provide parsimonious accounts of performance in single paradigms but to show generality as well.

In a preliminary attempt to evaluate whether the exemplar-generalization model provided better accounts than did the mixed-prototype model of our Experiment 3 data simply because of flexibility, we conducted power analyses involving the models. Specifically, we simulated how hypothetical subjects would classify stimuli assuming that their behavior followed a mixed-prototype model. We then tested the ability of the competing models to fit these simulated data. Because the focus was on performance at the late stages of learning, in these simulations we set the exemplar-memory parameter at $e = 1$ and the guessing parameter at $g = 0$. Furthermore, we simulated a broad-based prototype-abstraction process by assuming that individuals divided their attention weight equally among the six stimulus dimensions. We conducted separate series of simulations in which the sensitivity parameter was set at $c = 3$, $c = 6$, and $c = 12$. In each series of simulations, we generated 40 separate data sets in which a simulated observer classified each test instance four times (which was the number of classifications in our Experiment 3). We then fitted the mixed-prototype and exemplar-generalization models to these simulated data by using a maximum-likelihood criterion. When $c = 3$, the mixed-prototype model provided a better fit than did the exemplar-generalization for 39 of the 40 simulated data sets, and its mean log-likelihood fit to the data was significantly better than that of the exemplar-generalization model, $t(39) = 15.10$, $p < .001$. When $c = 6$, the mixed-prototype model fitted 37 of the 40 simulated data sets better than the exemplar model, with its mean log-likelihood fit again being significantly better, $t(39) = 7.69$, $p < .001$. At the highest level of sensitivity, the quantitative fit differences between the models grew smaller, but even here the mixed-prototype model was the correctly recovered model, providing better fits to 31 of the 40 data sets and yielding a significantly lower mean value of the log-likelihood statistic, $t(39) = 5.15$, $p < .001$. In general, the exemplar model was unable to fit these simulated data because it cannot simultaneously predict that observers will classify all of the training instances with high accuracy while also classifying the neighbors of the exceptions into the "prototype" category.

In summary, these preliminary analyses suggest that the exemplar-generalization model did not fit our Experiment 3 data better than did the mixed-prototype model simply because it is a more flexible model. If the mixed-prototype process truly governed performance, our power-analysis results suggest that the quantitative model-fit comparison should have revealed that fact.

Psychological Processes and the Context Model

In the present experiments, the context model achieved good quantitative fits to the individual subject classification data observed at both early and late stages of category learning. The results are consistent with the following ideas formalized in the model. At early stages of category learning involving highly separable-dimension stimuli, most observers attend selectively to single dimensions. As learning proceeds, if no single dimension provides adequate information to allow satisfactory performance, then observers spread attention to multiple dimensions of the objects. Thus, at early stages of learning, the exemplars stored in memory may be composed of only single-dimension or other low-dimension combinations of information, whereas in later stages of learning, the exemplars are composed of multiple-dimension combinations of information. In both cases, observers classify objects on the basis of their similarity to these stored exemplars. Furthermore, the classification decision-making process involves the use of response rules that are more deterministic than is predicted by a probability-matching rule (see also Estes, 1995). The EBRW model developed by Nosofsky and Palmeri (1997) provides a process-oriented account of how similarity-based comparisons lead to the retrieval of category exemplars and how this retrieved evidence drives classification decisions.

Because the process that we envision involves attention to single dimensions at the start of learning, an alternative interpretation of our data is that observers form rules as a basis for classification and later supplement these rules with exemplars or other sources of information. Indeed, a variety of multiple-system models that include rule formation as an important component have provided successful accounts of classification performance (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Johansen & Palmeri, in press; Nosofsky, Palmeri, & McKinley, 1994). The present research was not designed to distinguish between the predictions from multiple-system models involving rule formation and the predictions from the context model, and we believe that both approaches provide viable accounts of the present data. (For more extensive discussion of comparisons between multiple-system models and the context model, see Nosofsky and Johansen, 2000.) Rather, the research was intended to compare and contrast the predictions from the context model, with those from the mixed-prototype model of Smith et al. (1997; Minda & Smith, 2001; Smith & Minda, 1998, 2000). In our view, the present results provide compelling evidence against the psychological processes formalized in the mixed-prototype model.

Extensions of the Mixed-Prototype Model

It is unclear to us how to modify the mixed-prototype model to allow it to account for the generalization results from our Experiment 3 without fundamentally altering the character of this model. One possibility is that instead of assuming that observers store whole exemplars in memory, one can assume that they store parts of the exemplars. If a new item perfectly matches the stored exemplar on these parts, then it will be classified into the same category as the old instance, otherwise the prototype process is used. This mechanism is the one used in Nosofsky et al.'s (1994) proposed rule-plus-exception model of classification.

Regardless of the type of generalization mechanism that one might add to the exemplar-memory component of the mixed-prototype model, the key issue is as follows. According to the context model, a single representational system based on stored exemplars is assumed to mediate classification judgments. By contrast, according to the mixed-prototype model, two separate representational systems are involved, one based on prototypes and the other based on stored exemplars. The key question is whether one *needs* to extend the single-representation system assumed in the context model with the multiple-representation system assumed in the mixed-prototype model. Although Smith et al. (1997; Smith & Minda, 1998) have argued strongly that there is indeed evidence for prototype-abstraction mechanisms in classification, we have tried to show in this article that the standard, single-system exemplar-generalization model provides a natural account of all of their results. Furthermore, we demonstrated that the standard exemplar-generalization model provides a dramatically better account of patterns of generalization in classification than does the current version of the mixed model that assumes an all-or-none exemplar-memory process. The idea that observers abstract prototypes to represent categories is highly plausible. However, on grounds of parsimony, it is important for multiple-system theorists to provide clear and definitive evidence for the operation of prototype abstraction in people's category representations.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481.

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.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140.

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

Estes, W. K. (1995). Response processes in cognitive models. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 51–71). Hillsdale, NJ: Erlbaum.

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

Homa, D. (1984). On the nature of categories. *Psychology of Learning and Motivation, 18,* 49–94.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418–439.

Johansen, M. K., & Palmeri, T. J. (in press). Representational shifts in category learning. *Cognitive Psychology.*

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

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology, 71,* 331–338.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing.* Hillsdale, NJ: Erlbaum.

Lovett, M. (1998). Choice. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 255–296). Mahwah, NJ: Erlbaum.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–190). New York: Wiley.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision-bound and exemplar models of classification. *Perception & Psychophysics, 53,* 49–70.

McKinley, S. C., & Nosofsky, R. M. (1995). Investigations of exemplar and decision-bound models in large-size, 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, 5,* 207–238.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 241–253.

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.

Myung, I. J., Forster, M. R., & Browne, M. W. (Eds.). (2000). Model selection [Special issue]. *Journal of Mathematical Psychology, 44,* 190–204.

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,* 87–109.

Nosofsky, R. M. (1991a). 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. (1991b). Typicality in logically-defined categories: Exemplar similarity versus rule instantiation. *Memory & Cognition, 19,* 131–150.

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

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

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-

exception model of classification learning. *Psychological Review, 101,* 53–79.
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.
Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 736–753.
Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325–345.
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., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 659–680.
Stanton, R. D., Nosofsky, R. M., & Zaki, S. R. (in press). Comparisons between exemplar-similarity and mixed-prototype models of classification using a linearly separable structure. *Memory and Cognition.*
Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research.* New York: Wiley.
Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology.* San Francisco: Freeman.

Appendix

Proof of the Formal Identity Between the Extended Additive-Prototype Model and the Multiplicative-Prototype Model

According to the extended version of the additive-prototype model (Equation 11), in the absence of guessing,

$$P(\mathrm{A}|i) = \exp(\varphi \sum \delta_m w_m)/\{\exp[\varphi \sum \delta_m w_m] + \exp[\varphi \sum (1\text{-}\delta_m) w_m]\}$$
$$= 1/\{1 + \exp[\sum (1\text{-}2\delta_m)\varphi w_m]\}, \quad \text{(A1)}$$

where δ_m is as defined in the text (see Equation 4).

According to the multiplicative-prototype model (Equations 6–8), in the absence of guessing,

$$P(\mathrm{A}|i) = \exp(-c \sum w_m|x_{im} - P_{Am}|)/[\exp(-c \sum w_m|x_{im} - P_{Am}|)$$
$$+ \exp(-c \sum w_m|x_{im} - P_{Bm}|)]$$
$$= 1/\{1 + \exp[\sum (|x_{im} - P_{Am}| - |x_{im} - P_{Bm}|)\, c w_m]\}. \quad \text{(A2)}$$

Recall that the dimensions are binary valued and $P_{Am} \neq P_{Bm}$ for all m. Therefore, $|x_{im} - P_{Am}| - |x_{im} - P_{Bm}| = -1$ if item i matches P_A on dimension m, and equals $+1$ if item i matches P_B on dimension m. Therefore, $|x_{im} - P_{Am}| - |x_{im} - P_{Bm}| = 1 - 2\delta_m$. Substituting into Equation A2, it is seen that Equations A1 and A2 are formally identical with $c = \varphi$.

Received April 20, 2001
Revision received March 6, 2002
Accepted March 6, 2002 ■