Exemplar Model Account of Inference Learning

John K. Kruschke, Mark K. Johansen and Nathaniel J. Blair
Indiana University, Bloomington

June 14, 1999. Comments welcome.

Yamauchi and Markman (1998) claimed that a particular exemplar model could not fit data from both classification and inference learning. We show that a minor generalization of the model's response function, which allows for different degrees of decisiveness but which does not alter the exemplar-similarity processing of the model, allows the exemplar model to fit their data. Thus their evidence is inconclusive, and the same type of representation might underlie both types of learning.

Yamauchi & Markman (1998) claimed that a particular exemplar based model could not account for their results from category learning experiments. Their experiments involved two types of learning: One type of learning was standard *classification* learning, in which people learned to provide a missing category label when presented with all the stimulus features. A second type of learning was *inference* learning, in which people learned to provide a missing feature when presented with all the other stimulus features and the category label. Because their exemplar model predictions could not fit the data, Yamauchi & Markman (1998) argued that (1) "these models seem to require a major modification to account for inference transfer data" (p. 143) and (2) classification learning and inference learning "require different strategies" and "give rise to the formation of distinct category representations" (p. 144).

In this commentary we show that an exemplar model can, in fact, fit the data when its response function is generalized slightly to allow for different degrees of decisiveness. The exemplar-based core of the model remains unchanged. The generalized response function is the same one that Yamauchi and Markman (1998) used to test the Rational model, and is the same generalization as has been used with exemplar models in several previous publications. Hence the evidence provided by Yamauchi and Markman (1998) is inconclusive, and the fit of the slightly modified exemplar model to their data may even be taken as support for the opposite of their conclusions; that is, exemplar models might indeed be adequate to account for both classification and inference learning, and both types of learning could give rise to the same type of representation. In what follows, we first describe the classification and inference learning tasks, then the exemplar-based model, and then the fit of the model to Yamauchi and Markman's (1998) data.

Classification and Inference Learning

Yamauchi & Markman (1998) had participants learn about simple geometric figures that varied in form (circle/triangle), size (large/small), hue (red/green), position (left/right), and class ("Set A"/"Set B"). Thus, stimuli varied on five binary features, four of which were geometric, and one of which was verbal. In *classification* learning, participants saw the geometric features and had to predict the correct classification. In *inference* learning, participants saw the classification and three of the four geometric features, and had to infer the correct value of the missing geometric feature.

The assignment of stimuli to classifications was based on a prototype structure, such that the Set A prototype had exactly the opposite feature values as the Set B prototype, and stimuli were assigned to Set A or Set B according to which prototype had more matching geometric features. The prototypes themselves were never shown in training. This structure is sometimes referred to as a "family resemblance" structure. Table 1 shows the abstract design of the categories and learning tasks. Each row of the Table shows an instance that could occur on a learning trial. The feature preceded by a question mark ("?") is the missing feature that the participant had to predict, based on the other four presented features.

The upper section of Table 1 shows the instances used in *classification* learning. In all these cases, it is the class that must be inferred from the four geometric features. This section of Table 1 also clearly reveals the prototype structure of the classifications: Class 1 consists of those stimuli that have more geometric feature values of 1 than of 0. Class 0 consists of those stimuli that have more geometric feature values of 0 than of 1.

The lower section of Table 1 shows the instances used in *inference* learning. In all these cases, the class label and three of the four geometric feature values are provided, and the re-

This research was supported in part by NIMH FIRST Award 1-R29-MH51572-01.

For helpful comments on a previous version of this article, we thank Andrew Cohen, Robert Nosofsky, Teresa Treat, Peter Wood and Safa Zaki.

Correspondence can be addressed to John K. Kruschke, Department of Psychology, 1101 E. 10th St., Indiana University, Bloomington IN 47405-7007, or via electronic mail to kruschke@indiana.edu. The first author's world wide web page is at http://www.indiana.edu/~kruschke/

maining geometric feature value must be inferred. Yamauchi & Markman (1998) selected these inference training cases such that the missing feature value could be perfectly predicted from the class label. Thus, the missing feature value was always consistent with the prototype.

After training on the classification instances, or on the inference instances, or on a complete mixture of both, participants were then tested in a transfer phase on all the patterns and on the untrained prototypes. In terms of the abstract structure, there were just four distinct types of test patterns: First, there were the (trained) classification instances, in which three of the four features are consistent with the prototype. Second, there were the (untrained) classification prototypes, in which all four of the features are consistent with the prototype. Third, there were the (trained) inference instances in which the missing feature is consistent with the prototype. Fourth, there were the (untrained) inference instances in which the missing feature is *inconsistent* with the prototype.

We focus on the results of Yamauchi and Markman's Experiment 1, because complete data were provided only for this experiment, and because the results of Experiment 2 were qualitatively similar to the results of Experiment 1. The results from the transfer phase of Experiment 1 are reproduced in their entirety in Tables 2 and 3, *but Yamauchi & Markman (1998) emphasized and analyzed only the means of the four types of test patterns, across three kinds of training*. These means are highlighted in bold face in Tables 2 and 3, and can be summarized as follows (cf. Yamauchi & Markman, 1998, p. 134). For the inference tests in which the missing feature was inconsistent with the prototype, participants responded with the prototypical value more often than with the non-prototypical value (last row of Table 3). This tendency was greatest for people who experienced inference training, and least for people who underwent classification training. Across transfer types, accuracy was better for those transfer types that matched the training experienced by the participant. Finally, performance on the (untrained) prototypes was very high for all training types (last row of Table 2). Yamauchi and Markman argued that the dependency of performance on type of training indicated that inference training encourages encoding of "family resemblance information," whereas classification training encourages "attention to exception feature values."

Table 1
Abstract structures of stimuli used for classification and inference learning.

Geometric Feature				
1 (Form)	2 (Size)	3 (Hue)	4 (Position)	Class (Set A/B)
Classification learning				
1	1	1	0	?1
1	1	0	1	?1
1	0	1	1	?1
0	1	1	1	?1
0	0	0	1	?0
0	0	1	0	?0
0	1	0	0	?0
1	0	0	0	?0
Inference learning				
?1	1	1	0	1
?1	1	0	1	1
?1	0	1	1	1
?0	0	0	1	0
?0	0	1	0	0
?0	1	0	0	0
1	?1	1	0	1
1	?1	0	1	1
0	?1	1	1	1
0	?0	0	1	0
0	?0	1	0	0
1	?0	0	0	0
1	1	?1	0	1
1	0	?1	1	1
0	1	?1	1	1
0	0	?0	1	0
0	1	?0	0	0
1	0	?0	0	0
1	1	0	?1	1
1	0	1	?1	1
0	1	1	?1	1
0	0	1	?0	0
0	1	0	?0	0
1	0	0	?0	0

Exemplar Based Model

The Generalized Context Model (GCM)

Yamauchi & Markman (1998) applied the Generalized Context Model (GCM, Medin & Schaffer, 1978; Nosofsky, 1986) to their data, to examine whether a well-established exemplar-based model could address data from both inference and classification training. In the GCM, each distinct training item gives rise to an exemplar in memory, which

Table 2
Classification transfer; proportion of prototype-consistent responses. Observed and γGCM means are plotted in the left panel of Figure 2.

	Inference training			Mixed training			Classification training		
Stimulus	Obs.	γGCM	Y&M GCM	Obs.	γGCM	Y&M GCM	Obs.	γGCM	Y&M GCM
Non-prototypes									
1,1,1,0,?1	.864	.896	.895	.850	.923	.920	1.000	.952	.958
1,1,0,1,?1	.636	.685	.701	.950	.850	.853	1.000	.889	.893
1,0,1,1,?1	.818	.801	.789	.850	.875	.866	.826	.849	.848
0,1,1,1,?1	.773	.708	.709	.900	.850	.846	1.000	.969	.955
0,0,0,1,?0	.909	.896	.895	1.000	.923	.920	.913	.952	.958
0,0,1,0,?0	.727	.685	.701	.750	.850	.853	.783	.889	.893
0,1,0,0,?0	.773	.801	.789	.900	.875	.866	.870	.849	.848
1,0,0,0,?0	.636	.708	.709	.800	.850	.846	.957	.969	.955
Mean:	**.767**	**.773**	.774	**.875**	**.875**	.871	**.919**	**.915**	.914
Prototypes									
1,1,1,1,?1	.955	.929	.907	.900	.927	.951	.957	.978	.978
0,0,0,0,?0	.955	.929	.907	.950	.927	.951	.957	.978	.978
Mean:	**.955**	**.929**	.907	**.925**	**.927**	.951	**.957**	**.978**	.978

contains the features of the item. When a stimulus is presented with a missing class label that must be inferred, the model first computes the similarity of the stimulus to each exemplar in memory (based on the available features in the stimulus). Then each exemplar "votes" to fill in the missing class label with its stored value for the class label. The strength of the exemplar's vote is determined by its similarity to the stimulus. Thus, an exemplar in memory that is highly similar to the stimulus casts a strong vote in favor of its value for the missing class label, but an exemplar that is only slightly similar to the stimulus casts a weak vote in favor of its value for the missing class label.

Formally, the similarity η_{ij} of stimulus i to exemplar j is given by

$$\eta_{ij} = \exp\left(-c \sum_{k \neq \text{class}} w_k \left| x_{ik} - x_{jk} \right| \right) \tag{1}$$

where c is a scaling constant called the *specificity*, k is an index for the features over which similarity is computed, w_k is a weight on each feature, x_{ik} is the value of feature k in stimulus i, and x_{jk} is the value of feature k in exemplar j. The idea underlying this formalization is that similarity drops off exponentially with the distance, in psychological feature space, between the stimulus and the exemplar (Shepard, 1987). The distance in Equation 1 is computed using a city-block metric, with each featural dimension weighted appropriately for the task.

The similarities are transformed into choice probabilities according to a ratio rule (Luce, 1959), such that a value is chosen with probability equal to the summed similarity of exemplars in favor of that value, relative to the total similarity across all exemplars. When the missing feature is the class label, the probability of choosing class $C1$, given stimulus i, is given formally as

$$P(C1|i) = \sum_{j \in C1} \eta_{ij} \bigg/ \left[\sum_{j \in C1} \eta_{ij} + \sum_{j \in C0} \eta_{ij} \right] \tag{2}$$

where the sum in the numerator (i.e., $\sum_{j \in C1}$) is taken over all exemplars that vote in favor of value $C1$; that is, all exemplars j that have value $C1$ as their class label.

Extension of GCM to Address Inference

The GCM was originally formulated only for predicting classification labels on the basis of other features. Yamauchi & Markman (1998) made a reasonable extension of the GCM so that it could be applied to the inference task. They simply assumed that the classification label was another binary-valued feature, like the other stimulus features, and that Equations 1 and 2 could be applied analogously when the missing feature was geometric instead of the class label.

For example, when a stimulus i is presented with its form missing but its class provided, its similarity to exemplar j is given by

$$\eta_{ij} = \exp\left(-c \sum_{k \neq \text{form}} w_k \left| x_{ik} - x_{jk} \right| \right) \tag{3}$$

where the sum is taken over size, hue, position and class (i.e., not including form). The probability of selecting form value

Table 3

Inference transfer; proportion of prototype-consistent responses. Observed and γGCM means are plotted in the left panel of Figure 2.

	Inference training				Mixed training				Classification training			
Stimulus	Obs.	γGCM	$\gamma = 1$ GCM	Y&M GCM	Obs.	γGCM	$\gamma = 1$ GCM	Y&M GCM	Obs.	γGCM	$\gamma = 1$ GCM	Y&M GCM
Consistent with prototype												
?1,1,1,0,1	.909	.929	.750	.918	.800	.977	.750	.902	.652	.713	.646	.718
?1,1,0,1,1	.909	.918	.750	.918	1.000	.976	.750	.902	.652	.901	.861	.788
?1,0,1,1,1	.955	.966	.750	.918	.850	.973	.750	.902	.826	.815	.761	.788
?0,0,0,1,0	.909	.929	.750	.918	.900	.977	.750	.902	.652	.713	.646	.718
?0,0,1,0,0	.818	.918	.750	.918	.900	.976	.750	.902	1.000	.901	.861	.788
?0,1,0,0,0	.955	.966	.750	.918	1.000	.973	.750	.902	.739	.815	.761	.788
1,?1,1,0,1	.955	.945	.750	.918	.900	.966	.750	.902	.696	.727	.660	.723
1,?1,0,1,1	.909	.937	.750	.918	.950	.964	.750	.902	.957	.903	.863	.791
0,?1,1,1,1	.909	.935	.750	.918	1.000	.953	.750	.902	.870	.789	.734	.785
0,?0,0,1,0	.909	.945	.750	.918	.900	.966	.750	.902	.696	.727	.660	.723
0,?0,1,0,0	.909	.937	.750	.918	.950	.964	.750	.902	.913	.903	.863	.791
1,?0,0,0,0	1.000	.935	.750	.918	1.000	.953	.750	.902	.870	.789	.734	.785
1,1,?1,0,1	1.000	.930	.750	.918	.900	.960	.750	.902	.783	.755	.689	.723
1,0,?1,1,1	1.000	.967	.750	.918	1.000	.956	.750	.902	.957	.835	.781	.791
0,1,?1,1,1	1.000	.916	.750	.918	.950	.948	.750	.902	.870	.808	.752	.785
0,0,?0,1,0	.909	.930	.750	.918	1.000	.960	.750	.902	.783	.755	.689	.723
0,1,?0,0,0	1.000	.967	.750	.918	1.000	.956	.750	.902	.913	.835	.781	.791
1,0,?0,0,0	.955	.916	.750	.918	.900	.948	.750	.902	.826	.808	.752	.785
1,1,0,?1,1	.955	.923	.750	.918	.950	.955	.750	.902	.957	.896	.855	.759
1,0,1,?1,1	.955	.968	.750	.918	1.000	.952	.750	.902	.739	.801	.743	.759
0,1,1,?1,1	.955	.920	.750	.918	1.000	.945	.750	.902	.696	.762	.702	.751
0,0,1,?0,0	.909	.923	.750	.918	.950	.955	.750	.902	.826	.896	.855	.759
0,1,0,?0,0	.955	.968	.750	.918	1.000	.952	.750	.902	.739	.801	.743	.759
1,0,0,?0,0	.864	.920	.750	.918	1.000	.945	.750	.902	.826	.762	.702	.751
Mean:	**.938**	**.938**	.750	.918	**.950**	**.960**	.750	.902	**.810**	**.809**	.754	.763
Inconsistent with prototype												
?0,1,1,1,1	.818	.843	.750	.919	.700	.646	.750	.902	.522	.639	.609	.790
?1,0,0,0,0	.773	.843	.750	.919	.550	.646	.750	.902	.478	.639	.609	.790
1,?0,1,1,1	.909	.887	.750	.919	.800	.759	.750	.902	.609	.659	.627	.793
0,?1,0,0,0	.864	.887	.750	.919	.700	.759	.750	.902	.696	.659	.627	.793
1,1,?0,1,1	.909	.846	.750	.919	.850	.795	.750	.902	.696	.699	.661	.793
0,0,?1,0,0	.864	.846	.750	.919	.750	.795	.750	.902	.739	.699	.661	.793
1,1,1,?0,1	.864	.855	.750	.919	.900	.814	.750	.902	.739	.596	.562	.760
0,0,0,?1,0	.864	.855	.750	.919	.800	.814	.750	.902	.652	.596	.562	.760
Mean:	**.858**	**.858**	.750	.919	**.756**	**.754**	.750	.902	**.641**	**.648**	.615	.784

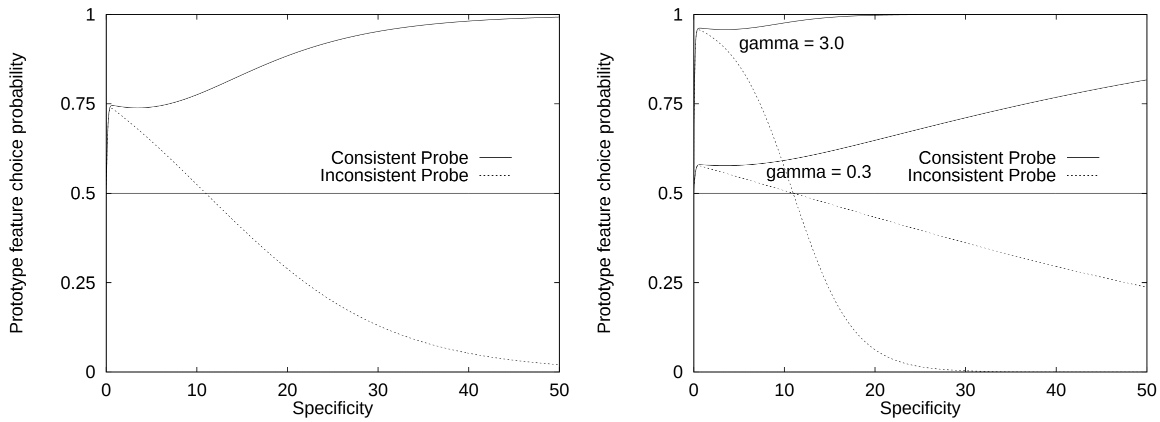

Figure 1. Left graph: Choice probability from Equation 4 for selecting the prototype-consistent feature, for consistent probe case ?11101 and inconsistent probe case ?01111. Right graph: Choice probability from Equation 5 for two different values of γ. (These graphs assume a classification label weight of 10.0 and a feature weight of 0.1.)

$F1$ is then given by

$$P(F1|i) = \sum_{j \in F1} \eta_{ij} \Bigg/ \left[\sum_{j \in F1} \eta_{ij} + \sum_{j \in F0} \eta_{ij} \right] \qquad (4)$$

where the sum in the numerator of Equation 4 is taken over all exemplars that vote in favor of value $F1$; that is, all exemplars that have value $F1$ as their form. Analogous equations apply for the other features of size, hue, and position.

Consider the predictions of this extension of the GCM for cases of inference transfer. Because the class label is positively correlated with all the geometric features, but the geometric features have zero correlation with each other, we assume that the weight on the class label is much larger than the weights on the geometric features. When c is extremely large, only the exactly matching exemplar in memory will have any influence on the choice. For very large c, therefore, the missing feature will be filled in essentially perfectly, whether it is consistent or inconsistent with the prototype. On the other hand, when c is small, only the class label will have any significant impact on the similarities, because the already small weights on the geometric features will be mitigated still further by being multiplied by a very small value of c. When c is small, therefore, all exemplars that match the class label of the stimulus will cast equally strong votes, and all exemplars that mismatch the class label of the stimulus will cast only negligible votes, and hence the choice proportion in favor of the prototype-consistent value will be .75, because three fourths of the exemplars vote in favor of the prototype-consistent value. Interpolating between these extremes, it can be seen that for intermediate values of c, the model predicts that missing features that are consistent with the prototype will be inferred with a choice proportion somewhere between 1.0 and 0.75, whereas missing features that are inconsistent with the prototype will be inferred to be consistent with the prototype with a choice proportion somewhere between 0.0 and 0.75. In other words, there must be some intermediate value of c such that Yamauchi and Markman's (1998) extension of the GCM makes qualitatively correct predictions for the inference cases.

To make the preceding discussion more concrete, consider two specific cases of inference transfer. One case is probe item ?11101, which matches a training item (11101) that has its first feature *consistent* with the prototype for category 1. The other case is probe item ?01111, which matches a training item (01111) that has its first feature *inconsistent* with the prototype for category 1. If we assume feature weights of the small value 0.1, and a classification label weight of the large value 10.0, then the probability of choosing the prototype-consistent value for the missing feature, as derived from Equations 3 and 4, is plotted in the left panel of Figure 1. For a range of specificity values, from about 0.5 to 11.0, the probability of choosing a prototype-*consistent* value for the missing feature of the *inconsistent* probe is greater than 0.5 and less than the choice probability for the consistent probe. This trend is qualitatively the same as that obtained from participants in the experiments conducted by Yamauchi & Markman (1998), as can be seen in Table 3.

Including Decisiveness in the Response Mapping

Yamauchi & Markman (1998) used a more general version of the choice probability function (i.e., Equation 2) when they tested a version of Anderson's (1990) Rational model of categorization. They used a version proposed by Nosofsky, Gluck, Palmeri, McKinley, & Glauthier (1994), which in turn was motivated by a version of the GCM proposed by Maddox and Ashby (1993, p. 54; see also Ashby & Maddox, 1993, Proposition 1). In this generalization of the choice probability function, there is greater flexibility in the extent to which a certain difference in similarities is mapped to a degree of choice preference. For example, suppose that for

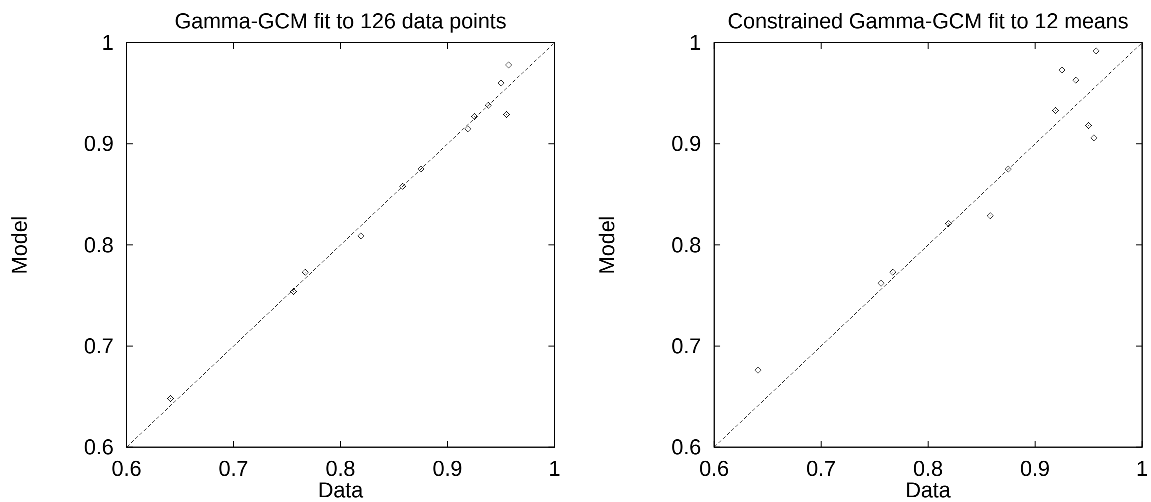

Figure 2. Left: Predictions by γ-GCM of 12 means, when fit to 126 cases using 33 parameters. Numerical data are provided in Tables 2 and 3. Right: Predictions by constrained γ-GCM, when fit directly to 12 means using 5 parameters. Numerical data are provided in Table 5.

a given stimulus, the summed similarity to exemplars of one category is 0.6, and the summed similarity to the exemplars of the other category is 0.4. Should this difference lead to an extremely decisive and strong choice preference, or should it lead instead to only a weak and indecisive preference? One way of parameterizing the degree of decisiveness is by raising the component summed similarities to a power, so that the largest summed similarity can be more or less dominant over the others. Formally, the mapping from summed similarity to choice proportion becomes

$$P(C1|i) = \left(\sum_{j\in C1} \eta_{ij}\right)^{\gamma} \Big/ \left[\left(\sum_{j\in C1} \eta_{ij}\right)^{\gamma} + \left(\sum_{j\in C0} \eta_{ij}\right)^{\gamma}\right] \tag{5}$$

where $\gamma > 0$. Notice that Equation 2 is a special case of this more general Equation 5 with $\gamma = 1$.

A psychological interpretation of the power γ is that it represents response confidence or *decisiveness*. When γ is large, then relatively small differences in summed similarity produce large response preferences, that is, more decisive responses, with probabilities closer to the extremes of 1 or 0. When γ is small, then response probabilities tend to be closer to 0.5. Maddox & Ashby (1993) provided an alternative interpretation in terms of noise in a decision criterion for a deterministic exemplar model. Nosofsky & Palmeri (1997, p. 291) described another interpretation of γ as a response threshold for a random walk model of speeded classification. Whereas neither of these interpretations might be directly applicable here, they do suggest that this mapping has several possible psychologically meaningful functions, including decisiveness.

As we demonstrate below, a value of γ greater than 1 is important for an adequate fit of the model. (Maddox & Ashby (1993) also found that good fits to their data demanded values of γ greater than 1.) When $\gamma = 1$, as in the model fits of Yamauchi & Markman (1998), the GCM can generate predictions that are qualitatively correct, as explained above, but that have absolute magnitudes that are off the mark. In other words, the predictions and the data covary appropriately, but the absolute deviation between predictions and data is large.

On the other hand, by incorporating the power γ in the response probability function (as in Equation 5), the absolute magnitude of the response probabilities is disentangled from the mechanism of summed similarities to exemplars. A larger value of γ, reflecting a higher decisiveness, permits a monotonic increase in the extremeness of the predicted response probabilities, without affecting the exemplar similarities. The right panel of Figure 1 illustrates how different values of γ affect choice probabilities, for the two cases of inference described earlier. As γ gets larger, the choice probabilities diverge more rapidly to the extremes; as γ gets smaller, the choice probabilities loiter closer to chance. The version of the GCM that allows γ to be freely estimated will be referred to henceforth as γGCM.

Alternative Formalization of Decisiveness

An alternative formalization of decisiveness has also been frequently used in the literature (e.g., Estes, 1988, 1994; Gluck & Bower, 1988; Kruschke, 1992). Instead of raising each component of summed similarity to a power, each component of summed similarity is exponentiated. Thus, instead of mapping summed similarities to choice probabilities via

Table 4
Best fitting parameter values and minimized error for the γGCM, and the minimized error for Yamauchi and Markman's (1998) application of the GCM.

Type of Training	w_{form}	w_{size}	w_{hue}	$w_{pos.}$	w_{class}	γ	RMSD	SSD	$\gamma=1$ SSD	Y&M SSD
					Classification transfer					
Inference	0.31	0.21	0.33	0.78	n/a	5.58	0.041	0.017	0.021	0.021
Mixed	5.10	4.01	5.10	2.08	n/a	0.30	0.062	0.039	0.040	0.040
Classification	1.03	32.23	2.71	1.59	n/a	0.78	0.055	0.030	0.031	0.031
					Inference transfer					
Inference	0.28	0.77	0.30	0.37	37.20	2.63	0.040	0.05	1.00	0.10
Mixed	0.30	0.49	0.62	0.73	0.21	22.52	0.063	0.13	1.12	0.38
Classification	0.51	0.71	1.39	0.22	36.80	1.32	0.084	0.22	0.31	0.48

the formula in Equation 5, we can use the mapping

$$P(C1|i) = \exp\left(\gamma \sum_{j\in C1} \eta_{ij}\right) \Big/ \left[\exp\left(\gamma \sum_{j\in C1} \eta_{ij}\right) + \exp\left(\gamma \sum_{j\in C0} \eta_{ij}\right)\right] \tag{6}$$

This mapping also has extensive precedents in the neural network and engineering literature, where it goes by the name of the "softmax" function (e.g., Bridle, 1990; Rumelhart, Durbin, Golden, & Chauvin, 1995). The basic motivation is the same; that is, to allow flexibility in the extent to which summed similarity is mapped to choice decisiveness.

The two formalizations of decisiveness can be seen to be closely related, after some algebraic manipulation. Notice first that Equation 5 can be re-expressed as a sigmoidal function (where $\text{sig}(x) = 1/[1+\exp(-\gamma x)]$):

$$P(C1|i) = 1 \Big/ \left[1+\exp\left(-\gamma\left(\log \sum_{j\in C1} \eta_{ij} - \log \sum_{j\in C0} \eta_{ij}\right)\right)\right] \tag{7}$$

Next, notice that Equation 6 can also be re-expressed as a sigmoidal function:

$$P(C1|i) = 1 \Big/ \left[1+\exp\left(-\gamma\left(\sum_{j\in C1} \eta_{ij} - \sum_{j\in C0} \eta_{ij}\right)\right)\right] \tag{8}$$

Thus, the two mappings are similar, but not identical, in that both are sigmoidal functions of the difference of summed similarities, but the first version takes the logarithm of the summed similarities before computing their difference.

In our simulations of Yamauchi and Markman's data, we initially used this alternative mapping (Equations 6 or 8), and found that it fit the data essentially as well the mapping in Equations 5 or 7. Thus, the specific formalization of decisiveness does not appear to be critical. In general, when the exemplar model is given the minor modification of monotonic flexibility in its response function, leaving its exemplar-similarity core untouched, then the model fits the data, as will be demonstrated below. In the remainder of the article we refer only to the first formalization of decisiveness (Equations 5 or 7), because this is the version used by Yamauchi & Markman (1998) for their simulations of the Rational model.

Fit of the Model to Data

Fit is Greatly Improved by Flexible Decisiveness

Tables 2 and 3 show the proportion of choices consistent with the prototype, as observed in human learners by Yamauchi & Markman (1998), as predicted by the best fitting γGCM, and as predicted by Yamauchi and Markman's (1998) application of the GCM. The values entered in the tables are the proportion of choices *consistent with the prototype value*. This proportion equals the proportion *correct* for items in Table 2 and for items in Table 3 that probed features consistent with the prototype. For the inference transfer items (Table 3) that probed a feature *inconsistent* with the prototype, however, this choice proportion is the proportion of *error*. For example, the inference transfer stimulus, ?01111, asks about the missing form of a item that was trained with form value 0. The correct response for this item would, therefore, be form value 0. The table shows, on the contrary, the proportion of choices of form value 1, which is consistent with the prototype.

Following the procedure used by Yamauchi & Markman (1998), the γGCM was fit to all 126 data points by fixing the specificity value at one ($c = 1$), and allowing the the attentional weights (w_k) and decisiveness (γ) to be freely estimated. Yamauchi and Markman (1998) also fixed $c = 1$ and allowed the attentional weights to be freely estimated, but effectively had the decisiveness fixed at a value of one. We tried to replicate the predictions of the GCM reported by Yamauchi & Markman (1998). Our best fits were concordant with those reported by Yamauchi and Markman (1998) for

Table 5
Predictions of constrained γGCM. (Also plotted in right panel of Figure 2.)

	Training					
	Inference		Mixed		Classification	
Stimulus Type	Obs.	γGCM	Obs.	γGCM	Obs.	γGCM
Classification Transfer						
Non-prototypes	.767	.773	.875	.875	.919	.933
Prototypes	.955	.906	.925	.973	.957	.992
Inference Transfer						
Consistent with Prototype	.938	.963	.950	.918	.810	.821
Inconsistent with Prototype	.858	.829	.756	.762	.641	.676

the classification transfer tests, but not for the inference transfer tests. We found that if we inappropriately used Equation 2 for all transfer items, including the inference transfer items, instead of the correct versions (such as Equation 4 for inferring missing form values, etc.), then we could reproduce the predictions reported by Yamauchi & Markman (1998) for the inference transfer tests.

It can be seen from Tables 2, 3 and 4 that the fit of the γGCM is much better than the fit of the GCM with $\gamma = 1$. This is especially evident for the inference transfer conditions (Table 3 and the lower half of Table 4). Even the worst fit of γGCM, inference transfer after classification training, provides a superb fit to the *mean* response proportions within types of transfer items. Specifically, examine in Table 3 the mean proportions for items consistent with the prototype, and for items inconsistent with the prototype. Across all three training conditions, it can be seen that for items inconsistent with the prototype (lower part of Table 3), the mean choice proportion is greater than .50 but less than the mean choice proportion for items consistent with the prototype (upper part of Table 3). Moreover, γGCM predicts these mean proportions extremely accurately, although it was fit to the individual item proportions. The left graph of Figure 2 shows the predicted means plotted against the empirical means. The correlation between the data and the predictions is .994, with 98.7% of the variance accounted for.

Fit to Essential Means

Although the fit to the individual test cases is greatly improved by letting decisiveness be a free parameter, Table 4 shows that the best fitting parameter values varied dramatically and uninterpretably from one condition to another. This does not necessarily indicate a shortcoming of the model, however. In fact, a wide range of parameter values fits the 12 means very well. When attempting to fit the probabilities of the 126 individual test cases, however, the parameter values are distorted to extremes to accommodate small amounts of residual variation. We fit all 126 individual test cases in the previous section only to replicate the procedure of Yamauchi & Markman (1998). In their data analysis, however, they described only the differences among the 12 means, which form the essence of the empirical results. Therefore we conducted a second fit, directly to the 12 means, using a constrained version of γGCM.

In the constrained γGCM, we allowed only five free parameters. First, we allowed only one weight for all the geometric features, across all conditions. This is motivated by the abstract category structure, which was symmetric with respect to geometric features. Any differences in feature weights would only reflect differences in the saliences of the physical instantiations, which we hoped were relatively small. We allowed a second weight for the classification label, but again fixed across conditions. This second weight is again motivated by the abstract structure of the categories, because the classification label is correlated with all of the geometric features, but no geometric feature is correlated with another. We provided the model with three values of decisiveness, γ, depending on the relationship between the transfer task and the training task. If the transfer task was the same as the training task (either both classification or both inference), then decisiveness should, presumably, be relatively high. If the transfer task differed from the training task (one classification and the other inference), then decisiveness should presumably be relatively low, because participants were attempting to make decisions that they had not previously been trained on. If the transfer task followed mixed training, then decisiveness might be yet another value, because mixed training entailed learning about the most cases.

Table 5 shows the predictions of the best fitting parameter values, and the right panel of Figure 2 shows the predicted means plotted against the empirical means. The correlation is .955, with 91.3% of the variance accounted for. The best-fitting parameter values yielded RMSD = 0.029, for a geometric feature weight of 0.599, a classification label weight of 1.90, a decisiveness of $\gamma = 4.06$ for the conditions with the transfer task the same as the training task, a decisiveness of $\gamma = 3.00$ for the mixed training conditions, and a decisiveness of $\gamma = 1.90$ for the conditions with the transfer task different from the training task. The ordinal relations between parameter values are as anticipated and psychologically meaningful.

Conclusion

The γGCM fits the means of the inference and classification transfer data very well, with parameter values that reflect intuitively plausible and theoretically meaningful differences in feature weights and decisiveness. The ability of the γGCM to fit the classification and inference data supports the notion that a common representation underlies both classification and inference knowledge. Whereas Yamauchi and Markman's (1998) hypothesis of different strategies and representations might be correct, the evidence presented in their article does not rule out the opposite hypothesis.

References

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

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

Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman-Soulié & J. Hérault (Eds.), *Neurocomputing: Algorithms, Architectures and Applications*, pp. 227–236. New York: Springer-Verlag.

Estes, W. K. (1988). Toward a framework for combining connectionist and symbol-processing models. *Journal of Memory and Language*, *27*, 196–212.

Estes, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

Gluck, M. A. & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, *27*, 166–195.

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

Luce, R. D. (1959). *Individual Choice Behavior*. New York: John Wiley.

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

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

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

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*, 352–369.

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

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In Y. Chauvin & D. E. Rumelhart (Eds.), *Backpropagation: Theory, architectures, and applications*, chap. 1, pp. 1–34. Hillsdale, NJ: Erlbaum.

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

Yamauchi, T. & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*, 124–148.