Research Report

RELATION BETWEEN THE RATIONAL MODEL AND THE CONTEXT MODEL OF CATEGORIZATION

Robert M. Nosofsky
Indiana University

Abstract—*A formal proof is provided that Anderson's (1990) rational model of categorization generalizes the Medin and Schaffer (1978) context model. According to the context model, people represent categories by storing individual exemplars in memory. According to the rational model, people represent categories in terms of multiple exemplar-clusters or prototypes. In both models, a multiplicative rule is used to compute the similarity of an item to the underlying category representations. In certain special cases, each multiple prototype in the rational model corresponds to an individual exemplar, and in these cases the rational model reduces to the context model. Preliminary quantitative comparisons between the models are illustrated to test whether the multiple-prototype view adds significant explanatory power over the pure exemplar view.*

Anderson (1990) recently proposed an important new model of categorization known as the *rational model*. The purpose of the present note is to consider formal relations between the rational model and Medin and Schaffer's (1978) well-known *context model* of categorization. The main point is to prove that a slightly modified version of Anderson's (1990) rational model generalizes the context model. Relations between the rational model and the *fuzzy logical model of perception* (Massaro, 1987; Massaro & Friedman, 1990; Oden & Massaro, 1978) are also discussed.

The proof that the rational model generalizes the context model has important implications for future research strategies aimed at comparing these two models. For example, Anderson (1990) demonstrated that the rational model could account for numerous important categorization phenomena, such as prototype effects, effects of similarity to specific training exemplars, sensitivity to correlated features, and probability matching behavior. However, the context model has been demonstrated previously to account for all these major phenomena (Busemeyer, Dewey, & Medin, 1984; Estes, 1986; Medin & Schaffer, 1978; Medin, Altom, Edelson, & Freko, 1982; Nosofsky, 1988a, 1990). In light of the finding that the rational model generalizes the context model, the rational model's success in the aforementioned domains is unsurprising: By necessity, a more general model must account for data at least as well as its special cases.

The critical issue is whether the extra free parameters accorded the rational model lead to clear improvements in explanatory power over its special case, the context model. Otherwise, on grounds of parsimony, the simpler context model is to be preferred. I illustrate some preliminary explorations of this issue in this note.

AN INTUITIVE DESCRIPTION OF THE RATIONAL MODEL, THE CONTEXT MODEL, AND THEIR FORMAL RELATION

According to the context model, people store the individual training exemplars of categories in memory, and make classification decisions on the basis of the summed similarity of items to the exemplars of the alternative categories. Overall similarity between exemplars is computed by using a multiplicative-combination rule, in which similarities along individual component dimensions are multiplied together.

In Anderson's (1990) rational model, individual exemplars become grouped together into clusters during the learning process. The probability that a particular exemplar is grouped into a cluster depends on how similar the exemplar is to the cluster's central tendency, and on the *prior probability* that items are grouped into the cluster. This prior probability is determined jointly by the current size of each cluster and the value of a *coupling* parameter, which is a free parameter in the model. The model also has mechanisms for computing the probability with which membership in each cluster signals a given category label. According to the model, when classification decisions are to be made, the observer computes the similarity of the item to the central tendency of each cluster, and sums these similarities weighted by the respective category-label probabilities of each cluster. Similarity to the central tendency of each cluster is computed by using a multiplicative-combination rule that is isomorphic to the one used in the context model.

The key intuition is that when the value of the coupling parameter is zero, each individual exemplar forms its own cluster, as in the context model. In this special case, the rational model reduces to the context model. (Technically, as explained later, it is a slightly modified version of the rational model that reduces to the context model when the coupling parameter is set at zero.) The rational model also generalizes the context model by allowing for similarities between the category labels that are assigned to stimuli. However, the major conceptual generalization that is involved is the clustering idea that results from the action of the coupling parameter.

The clustering idea is interesting, because it provides an elegant formalization of the hypothesis that people store multiple prototypes in memory. In much previous work, Medin and his associates (Medin, Dewey, & Murphy, 1983; Medin & Schaffer, 1978; Medin & Smith, 1981) and Nosofsky (1987, 1988b, 1991b) demonstrated that the exemplar-based context model consistently outperforms prototype models in quantitatively predicting classification performance. A natural idea is that instead of literally storing

Address correspondence to: Robert Nosofsky, Department of Psychology, Indiana University, Bloomington, IN 47405; e-mail: nosofsky@ucs.indiana.edu.

each individual exemplar as a unique memory trace, and instead of simply storing a single category prototype, the cognitive system might do something intermediate, namely store multiple prototypes to represent a category. What was lacking in previous work, however, was a principled model for predicting which multiple prototypes are formed. The rational model provides such an account, in that the multiple clusters act as multiple prototypes.

Another special case of the rational model is also of interest. When the value of the coupling parameter is unity, and there is zero similarity among category labels, exemplars with a given category label are all grouped together in one grand cluster. In other words, the clusters that are formed correspond to the experimentally defined categories. In this case, the rational model takes the form of a multiplicative-similarity prototype model (Estes, 1986; Golden & Rumelhart, 1989; Medin, Altom, & Murphy, 1984; Nosofsky, in press), important examples of which are Massaro's (1987) fuzzy logical model of perception and Gluck and Bower's (1988) (nonconfigural) adaptive network models. Details are provided by Nosofsky (1991a, in press).

FORMAL STATEMENT

Assume that the stimuli vary along M binary-valued dimensions. The Medin and Schaffer (1978) context model has M free similarity parameters, one similarity parameter (s_m) for each of the M dimensions. Anderson's (1990) rational model has M + 2 free parameters, of which M free sensitivity parameters (α_m) are analogous to the dimensional-similarity parameters in the context model. There is also a category-label sensitivity parameter (α_{M+1}), reflecting the similarity among the alternative category labels. Finally, there is a coupling parameter (c), which influences the degree to which individual exemplars are grouped into clusters during the learning process.

The rational model is formally identical to the context model when the value of the category-label sensitivity parameter, α_{M+1}, is equal to 0 (zero similarity among category labels); the value of the coupling parameter c during the cluster-formation process is zero; and the value of the coupling parameter during the classification-decision process is unity. A proof is provided in the Appendix. The idea that the coupling parameter can take on different values during the cluster-formation process and the classification-decision process represents a departure from the assumptions introduced by Anderson (1990), but I argue in the Appendix that the idea is reasonable. Basically, by setting the coupling parameter at a low value during the cluster-formation process, the observer establishes fine-grained clusters of category information in memory. By switching the coupling parameter to a high value at time of decision, the observer makes full use of this fine-grained information.

PRELIMINARY MODEL COMPARISONS

Because the rational model generalizes the context model, it is of interest to explore whether the added parameters contribute in significant ways to improving the fits of the context model, or whether the simpler model will suffice. In many situations we would expect the predictions of the models to be extremely similar. If a group of exemplars is clustered closely together in a multidimensional space, then the summed similarity of a test item to those exemplars can be well approximated by computing the similarity of the test item to the cluster's central tendency and then multiplying by the number of objects in the cluster. This computation is essentially what is performed by the rational model—it groups similar exemplars into clusters, computes the similarity of an item to the cluster's central tendency, and multiplies by the frequency with which exemplars of the cluster have been experienced (the cluster's "prior"). Exemplars that are dissimilar to those in one cluster form their own separate clusters.

In a preliminary comparison of the models, I fitted them to 11 sets of classification transfer data reported previously by Medin and his associates (see Table 1). All experiments involved fairly standard classification learning and transfer paradigms, in which people learned to classify the items on the basis of induction over individually presented exemplars. In all cases, there were sixteen transfer stimuli varying along four binary-valued dimensions, and subjects were to classify them into two categories.

In fitting the rational model, 100 simulations were conducted in which the training exemplars were presented in a random order. (The clusters that are formed by the rational model vary depending on the ordering of training exemplars.) The results of each simulation were then averaged to generate the predictions of the model. (The same 100 random sequences were used in all simulations.) Because a "hill-climbing" search procedure was used to find the best-fitting parameters, the results of the model-fitting analyses should be interpreted with caution. Unfortunately, the clusters that are formed by the rational model vary depending on the values of the parameters in the model. With only slight changes in parameter values, large changes in predictions can occur because new clusters are formed. Thus, the possibility of the search routine getting stuck in local minima is especially troublesome. To guard against this possibility, several different starting configurations were used in each parameter search, but local minima may still remain.

The various models were fitted to the sets of classification transfer data by minimizing the sum of squared deviations between predicted and observed response probabilities. In Table 1, column 1 gives the results for the context model, i.e., the rational model with c = 0 during the cluster-formation process, α_{M+1} = 0, and c = 1 during the classification-decision process. Column 2 gives the results for the rational model with c allowed to be a free parameter during cluster formation. As is assumed for the context model, however, we assume α_{M+1} = 0 and c = 1 during classification decisions. To make sure that setting c = 1 during the classification-decision process does not adversely affect the predictions of the rational model, column 3 gives the results when c is allowed to be a free parameter, but is held fixed across the cluster-formation and classification-decision phases. Finally, column 4 gives the fits for the rational model with α_{M+1} a free parameter to allow for the potential role of category-label similarity.

Table 1. *Sum of squared deviations between predicted and observed category 1 response probabilities for the alternative versions of the rational model*

Data Set	Version 1	2	3	4
Medin & Schaffer (1978)				
Exp. 2	.060	.043	.028	.031
Exp. 3	.031	.031	.060	.029
Exp. 4	.078	.060	.162	.055
Medin & Smith (1981)				
Proto. Instruct.	.029	.028	.063	.026
Rule Instruct.	.049	.049	.053	.044
Medin, Dewey, & Murphy (1983)				
Last-name infinite	.043	.037	.037	.037
Last-name only	.053	.047	.059	.046
Medin, Altom, & Murphy (1984)				
Exp. 1 (examples only)	.052	.052	.057	.048
Exp. 2 (examples only, geometric)	.041	.032	.053	.024
Exp. 3 (examples only, verbal)	.072	.062	.090	.048
Medin, Altom, Edelson, & Freko (1982)				
Exp. 4	.015	.015	.018	.015

Note. Version of rational model:
1. c = 0 during cluster formation, c = 1 during classification decision, $\alpha_{M+1} = 0$ (i.e., the context model).
2. c free to vary during cluster formation, c = 1 during classification decision, $\alpha_{M+1} = 0$.
3. c free to vary during cluster formation, held fixed at same value during classification decision, $\alpha_{M+1} = 0$.
4. c free to vary during cluster formation, c = 1 during classification decision, α_{M+1} free to vary.

Comparing the fits in column 2 with those in column 1, we see little evidence that adding the coupling parameter markedly improves the fit of the context model. Furthermore, because the fits in column 3 are no better (and are sometimes substantially worse) than those in column 2, the assumption that c = 1 during the classification-decision phase does not adversely affect the rational model's predictions. Finally, there is also little evidence that similarity between category labels played a major role in these experiments, because the fits in column 4 are essentially the same as those in columns 1 and 2. Because the category labels were highly distinctive in Medin's experiments, and these features were the ones that subjects were explicitly trying to predict, it is not surprising that the category-label similarity parameter does not contribute to the model fits.

In summary, at least in the standard classification learning/transfer paradigms conducted by Medin and his colleagues, elaborating the context model by adding a coupling parameter and a category-label similarity parameter does not appear to substantially improve its fits to the data. This result is not meant to imply, however, that there is no advantage to the rational model. Anderson (1990), for example, reports that the rational model is able to predict an inverse bate-rate effect that was observed in a recent study by Medin and Edelson (1988), something that the standard context model is unable to do. Likewise, Heit (1990) reports that the rational model successfully predicts indirect, mediated generalization phenomena, which the standard context model is unable to do. (Other elaborated versions of the context model have also successfully predicted these phenomena, however—see Medin & Edelson, 1988, and Heit, 1990). Progress in future research might be facilitated by systematic comparisons between the rational model and the context model, and understanding the conditions in which the coupling parameter plays a significant role in explaining psychological phenomena.

Finally, extensive research involving the context model during the past decade may also help guide future developments of the rational model. Nosofsky (1984, 1986) demonstrated that the context model is closely related to the classic theories of categorization and similarity developed by Shepard and his colleagues (Shepard, 1958, 1964, 1987; Shepard & Chang, 1963; Shepard, Hovland, & Jenkins, 1961). For example, the multiplicative-similarity rule that is assumed in the context model arises if psychological distance between exemplars is described by a city-block metric, and similarity is an exponential decay function of distance in psychological space (Shepard, 1987). This relation led Nosofsky (1984, 1986) to propose the *generalized context model* (GCM), in which exemplars are represented as points in psychological scaling solutions, with similarity between exemplars being a decreasing function of their distance in the space. This generalization allowed the context model to be applied in a straightforward way to predict categorization in continuous-dimension domains, in addition to discrete ones. Anderson (in press) recently proposed an extension of the rational model for application in continuous-dimension domains, although the assumptions for modeling similarity differ somewhat from those used in the GCM. In addition to testing for the importance of the coupling parameter and the clustering process, future research will need to test between the similarity assumptions embodied in the continuous-dimension versions of the context model and the rational model.

Acknowledgments—This work was supported by grants BNS 87-19938 from the National Science Foundation and PHS R01 MH48494-01 from the National Institute of Mental Health to Indiana University. My thanks to John Anderson and Jerry Busemeyer for their comments on an earlier version of this article, and to Jerry Busemeyer for helpful discussions.

REFERENCES

Anderson, J.R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R. (in press). The adaptive nature of human categorization. *Psychological Review*.

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

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

Gluck, M.A., & Bower, G.H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227–247.

Golden, R.M., & Rumelhart, D.E. (1989). *A relationship between the context model of classification and a connectionist learning rule.* Unpublished manuscript.

Heit, E. (1990). *Reasoning from examples.* Unpublished Ph.D. dissertation, Stanford University.

Massaro, D.W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D.W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97*, 225–252.

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

Medin, D.L., Altom, M.W., & Murphy, T.D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 3*, 333–352.

Medin, D.L., Dewey, G.I., & Murphy, T.D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 607–625.

Medin, D.L., & Edelson, S.M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117*, 68–85.

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

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

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, & Cognition, 13*, 87–109.

Nosofsky, R.M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700–708.

Nosofsky, R.M. (1988b). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 54–65.

Nosofsky, R.M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34*, 393–418.

Nosofsky, R.M. (1991a). Relation between the rational model and the context model of categorization. *Cognitive Science Report #39, Indiana University.*

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

Nosofsky, R.M. (in press). Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *Essays in Honor of W.K. Estes.*

Oden, G.C., & Massaro, D.W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*, 172–191.

Shepard, R.N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology, 55*, 509–523.

Shepard, R.N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*, 54–87.

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

Shepard, R.N., & Chang, J.J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology, 65*, 94–102.

Shepard, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75* (13, Whole No. 517).

(Received 3/4/91; Accepted 5/17/91)

APPENDIX

Formal Description of the Context Model, the Rational Model, and a Proof of their Relation

The following development is limited to situations involving stimuli varying along M binary-valued dimensions. The stimuli are to be classified into K categories. Let t denote a test stimulus, and let t_m denote the value of t on dimension m. Exemplars of category J are denoted by j, with the value of exemplar j on dimension m denoted by j_m.

According to Medin and Schaffer's (1978) context model, the probability that t is classified in category J is found by summing the similarity of t to all exemplars of category J, and then dividing by the summed similarity of t to all exemplars of all K categories. Similarity between exemplars t and j is computed by using the following multiplicative rule:

$$s(t,j) = \prod_m s_m^{\delta_m(t,j)} \qquad (1)$$

where s_m is a similarity parameter representing the intrinsic salience or "attention" devoted to dimension m, and $\delta_m(t,j)$ is an indicator variable equal to one if t and j mismatch on dimension m, and equal to zero if they match. Thus, according to the context model, the probability of classification response J given presentation of t is given by

$$P(R_J \mid t) = \frac{\sum_{j \in C_J} \prod_m s_m^{\delta_m(t,j)}}{\sum_K \sum_{k \in C_K} \prod_m s_m^{\delta_m(t,k)}}. \qquad (2)$$

I describe the rational model by first reviewing the algorithm for creating clusters. Given a particular clustering, I then review how the model is used to predict classification probabilities.

At any given point in the learning process, there is some clustering of the stimuli that have been presented so far. Upon presentation of stimulus t the subject computes the probability that the stimulus belongs to each cluster h, pclus(h | t), and also computes the probability that the stimulus comes from a completely new cluster, pclus(0 | t). The stimulus is then assigned to the cluster that has maximum probability. [Note that on the first trial of learning, before any clusters exist, a stimulus automatically forms its own new cluster. Also, in certain special circumstances, the cluster probabilities are undefined—e.g., see Nosofsky (1991a). In these circumstances, I assume that the stimulus automatically forms its own new cluster.]

The probability of cluster h given t is given as in Bayes's Theorem:

$$\text{pclus}(h \mid t) = \frac{p(t \mid h)\,\text{prior}(h)}{\sum_k p(t \mid k)\,\text{prior}(k)} \qquad (3)$$

where p(t | h) is the probability of the features composing stimulus t given membership in cluster h, and prior(h) is the prior probability of cluster h. The prior probability that t belongs to cluster h is given by

$$\text{prior}(h) = \frac{cn_h}{(1 - c) + cn} \qquad (4a)$$

where c is the coupling probability, n_h is the number of objects assigned to cluster h so far, and n is the total number of objects seen so far. The prior probability that t belongs to a new cluster (0) is given by

$$\text{prior}(0) = \frac{1 - c}{(1 - c) + cn}. \qquad (4b)$$

The conditional probability of the features composing t given membership in cluster h, p(t | h), is given by

$$p(t \mid h) = \prod_m p(t_m \mid h), \qquad (5)$$

where, as described by Anderson (1990, p. 143),

$$p(t_m \mid h) = \frac{n_h(t_m) + \alpha_m}{n_h + n(m) \cdot \alpha_m}. \qquad (6)$$

In Equation 6, $n_h(t_m)$ is the number of objects in cluster h with value t_m on dimension m, n_h is the total number of objects in cluster h, n(m) is the number of values on dimension m [n(m) = 2 in the present application], and α_m is a parameter that reflects sensitivity on dimension m. Smaller values of α_m reflect greater sensitivity. Following Anderson (1990, p. 116), I assume that in Equation 6 sensitivity is constant across all values within a dimension.

An important contribution made by Anderson (1990) is the formalization of the idea that the category label associated with each object is simply another feature. To the extent that category labels are similar, objects that are assigned by the experimenter to different category labels can nevertheless be grouped in the same cluster. The probability of category-label J given membership in cluster h, P(J | h), is calculated as in Equation 6. The category label can be viewed as the value on dimension M + 1.

Suppose now that stimulus t is presented (without its category label) and the subject is asked to supply the category label. According to the rational model, the probability that the subject classifies t as having category-label J is given by

$$P(R_J \mid t) = \sum_h \text{pclus}(h \mid t)\, P(J \mid h) \qquad (7)$$

where pclus(h | t) and P(J | h) are defined as before. The summation is over all clusters h, including the new cluster.

I now prove that the rational model just described reduces to the context model (Eq. 2) when the sensitivity parameter on the category-label dimension (α_{M+1}) is equal to zero, the value of the coupling parameter (c) during the cluster-formation process is set at zero, and the value of the coupling parameter during the classification-decision process is set at unity.

To begin, consider what happens during the cluster-formation process when the coupling parameter is set at zero. According to Equations 4a and 4b, the prior probability of t belonging to each pre-existing cluster h is zero, and the prior probability of t belonging to a new cluster is one. Substituting into Equation 3, we see that pclus(h | t) = 0 for all pre-existing clusters h, and pclus(0 | t) = 1. Thus, because t is assigned to the cluster with maximum probability, we see that each individual exemplar presentation leads to the formation of a new cluster. By the end of the cluster-formation process, each cluster consists of one individual exemplar.

Expanding the classification-decision rule of the rational model (Eq. 7) by substituting the expression for pclus(h | t) (Eq. 3) yields

$$P(R_J \mid t) = \sum_h \left[\frac{p(t \mid h)\,\text{prior}(h)}{\sum_k p(t \mid k)\,\text{prior}(k)} \cdot P(J \mid h) \right]. \qquad (8)$$

Assuming that the coupling parameter is set at unity during the classification-decision process, then the prior probability of a new cluster is zero, and the value prior(h) is the same (1/n) for all pre-existing clusters h, because each cluster consists of one exemplar (see Eqs. 4a and 4b). Thus, Equation 8 can be rewritten as

$$P(R_J \mid t) = \sum_h \left[\frac{p(t \mid h)}{\sum_k p(t \mid k)} \cdot P(J \mid h) \right] = \frac{\sum_h P(t \mid h)\, P(J \mid h)}{\sum_k p(t \mid k)} \qquad (9)$$

where the sum is over all (single-exemplar) clusters.

Because we assume that the sensitivity parameter on the category-label dimension (α_{M+1}) is zero, note that each P(J | h) equals either zero or one (see Eq. 6). If the exemplar defining cluster h has category-label J, then P(J | h) = 1, whereas if the exemplar defining cluster h does not have category-label J, then P(J | h) = 0. Thus, Equation 9 can be rewritten as

$$P(R_J \mid t) = \frac{\sum_{j \in C_J} p(t \mid j)}{\sum_K \sum_{k \in C_K} p(t \mid k)}. \qquad (10)$$

The index $j \in C_J$ denotes that we are summing over all clusters (exemplars) with category-label J. (Note that the denominator in Eq. 10 is the same as in Eq. 9. The partitioning of the clusters into categories is simply made explicit in the notation.)

Substituting the expression for p(t | j) (Eqs. 5 and 6) into Equation 10 yields

$$P(R_J \mid t) = \frac{\sum_{j \in C_J} \prod_{m=1}^{M} \left(\frac{n_j(t_m) + \alpha_m}{n_j + 2\alpha_m} \right)}{\sum_K \sum_{k \in C_K} \prod_{m=1}^{M} \left(\frac{n_k(t_m) + \alpha_m}{n_k + 2\alpha_m} \right)} \qquad (11)$$

Note that for all clusters, $n_j = 1$; $n_j(t_m) = 1$ if cluster (exemplar) j has value t_m

on dimension m; and $n_j(t_m) = 0$ if cluster (exemplar) j does not have value t_m on dimension m. Thus, we can rewrite Equation 11 as

$$P(R_J \mid t) = \frac{\sum_{j \in C_J} \prod_{m=1}^{M} \left(\frac{1+\alpha_m}{1+2\alpha_m}\right)^{1-\delta_m(t,j)} \left(\frac{\alpha_m}{1+2\alpha_m}\right)^{\delta_m(t,j)}}{\sum_K \sum_{k \in C_K} \prod_{m=1}^{M} \left(\frac{1+\alpha_m}{1+2\alpha_m}\right)^{1-\delta_m(t,k)} \left(\frac{\alpha_m}{1+2\alpha_m}\right)^{\delta_m(t,k)}}, \quad (12)$$

where $\delta_m(t,j) = 1$ if $t_m \neq j_m$, and $\delta_m(t,j) = 0$ if $t_m = j_m$. Dividing all terms by $\Pi (1 + \alpha_m)/(1 + 2\alpha_m)$ yields

$$P(R_J \mid t) = \frac{\sum_{j \in C_J} \prod_{m=1}^{M} \left(\frac{\alpha_m}{1+\alpha_m}\right)^{\delta_m(t,j)}}{\sum_K \sum_{k \in C_K} \prod_{m=1}^{M} \left(\frac{\alpha_m}{1+\alpha_m}\right)^{\delta_m(t,k)}}. \quad (13)$$

Comparing Equation 13 to Equation 2, we see that this version of the rational model is formally identical to the context model, with $s_m = \alpha_m/(1 + \alpha_m)$. When s_m in the context model equals 0, α_m in the rational model equals 0, and when s_m in the context model equals one, α_m in the rational model equals ∞.

As explained previously, setting $c = 1$ during the classification-decision process implies that the subject makes full use of the exemplar information stored in memory when making his or her classification decisions. When $c \neq 1$, the value pclus $(0 \mid t)$ would not be equal to zero (see Eqs. 3 and 4b). Thus, the subject would be giving some weight to the new cluster when making classification decisions (Eq. 7). In the extreme case in which $c = 0$ during classification decisions, all weight would be given to the new cluster, and none of the previously stored exemplar information would be used.

Anderson (1990, p. 113) previously discussed structural similarities between the rational model and the context model, and suggested that the context model is the rational model in the presence of poor category structure. This suggestion is difficult to evaluate because the condition "poor category structure" is not explicitly defined, and no formal proof was provided. It should be emphasized, however, that if *any* two exemplars in the set are similar, and the coupling probability is even modest, then the exemplars will combine in a cluster and the models will no longer be formally identical. Even if each individual exemplar forms its own cluster, the models are still not formally identical if the coupling parameter is not set at unity during the classification-decision process, because the new cluster would be given some weight in the Equation 7 decision rule. Finally, previous tests of the context have not incorporated the assumption that the category labels themselves contribute to overall similarity among exemplars, as is allowed in the rational model. All three ideas that are incorporated in the rational model—the possibility of an exemplar-clustering process, use of the new cluster in making decisions, and similarity between category labels—have some psychological plausibility, and may well give the rational model some advantages in predicting categorization data.