# Dimension-Wide vs. Exemplar-Specific Attention in Category Learning and Recognition

**Yasuaki Sakamoto (yasu@psy.utexas.edu)**
Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA
**Toshihiko Matsuka (matsuka@psychology.rutgers.edu)**
RUMBA, Department of Psychology, Rutgers University
Newark, NJ 071012 USA
**Bradley C. Love (love@psy.utexas.edu)**
Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA

## Abstract

Items that violate a category rule are remembered better than items that follow the rule. This finding cannot be predicted by exemplar models when all exemplars share the same attention along a dimension. With dimension-wide attention, violating and rule-following items are treated equally. When each exemplar selects which dimensions to attend to, exemplar models can predict the memory advantage for violating items. With exemplar-specific attention, attention is distributed uniformly for exemplars encoding violating items but is allocated to the rule dimension of exemplars encoding rule-following items. This differential attention makes violating items distinctive in memory. In addition to exemplar-specific attention, exemplar models need the ability to distinguish important errors from negligible ones to predict better memory for items that violate a stronger than a weaker rule.

Humans are confronted with more information than they can process. Consequently, the ability to selectively attend to salient information is fundamental to our cognitive behavior. Many category learning models utilize the same attention at all locations along a dimension in the representational space (e.g., Kruschke, 1992; Love, Medin, & Gureckis, 2004; Nosofsky, 1986). The dimension-wide attention is well suited for many artificial category learning studies, in which categories are symmetric and category members are differentiated by the values on the same dimensions. For example, category A members may be large on the size dimension, whereas category B members may be small. In natural categories, however, there are inconsistent items that do not follow the structure. For example, penguins do not fly but are members of category birds, whereas bats fly but belong to category mammals. Dimension-wide attention may not fare as well when categories contain inconsistent members.

Some laboratory work does suggest that attention is specific to the region along a dimension in the representational space (e.g., Aha & Goldstone, 1992; Barsalou & Medin, 1986; Lewandowsky, Kalish, & Ngang, 2002). Humans attend to different dimensions of an item depending on the context the item is in. For example, humans may attend to the color dimension when shopping for clothing but not as much when shopping for a computer.

Exemplar models have a long history of explaining key psychological phenomena in the category learning research (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). However, one finding that is problematic for exemplar models is that people better remember items that violate a structure (e.g., rule) than structure-consistent (e.g., rule-following) items (Palmeri & Nosofsky, 1995; Sakamoto & Love, in press). By storing every studied item as a separate trace and using the dimension-wide attention, current exemplar models treat consistent and inconsistent items in the same fashion and cannot predict the memory advantage for inconsistent items.

In this paper, we show that an exemplar model with exemplar-specific attention (see Kruschke, 2001 for a related model with exemplar-specific "specificity") can differentiate violating items from rule-following items and predict the memory advantage for violating items. While exemplars encoding rule-following items result in attention allocated to the rule dimension, exemplars encoding violating items result in attention distributed to the non-rule dimensions. This differential attention makes violating items distinctive in memory.

We further show that in addition to the exemplar-specific attention, exemplar models need a mechanism that accentuates larger errors and minimizes the impact of smaller ones to predict a better memory for items that violate a stronger rule than items that violate a weaker rule (e.g., Sakamoto & Love, in press). In the remainder of the paper, we review previous category learning work that examines recognition memory for violating items, introduce the models, present model fits to previous findings, and discuss our modeling results.

## Memory Advantage for Exceptions

Palmeri and Nosofsky (1995) found that humans remember exceptions to the category rule better than items that follow the rule. In their study, subjects learned to classify geometric stimuli into two contrasting categories. An imperfect rule successfully classified the majority of study items (e.g., most small items were in category A, whereas most large

items were in category B), but two exceptions violated the rule (e.g., a large item that was a member of category A). Following learning, subjects completed a recognition test consisting of studied items and novel items that served as foils. The basic finding was that people recognized the exceptions better than the studied rule-following items. The special status of violating items is also suggested by the schema (e.g., Rojahn & Pettigrew, 1992) research and is not specific to category learning studies.

Exemplar models with dimension-wide attention, such as ALCOVE (Kruschke, 1992) and the context model (Medin & Schaffer, 1978), cannot predict the memory advantage of exceptions. Exemplar models store every studied item as a separate trace. The likelihood of recognizing an item is determined by the item's absolute similarity to all exemplars from both categories A and B. Exemplar models cannot predict the enhanced memory for exceptions because the exceptions share the same similarity relations with other items in memory as rule-following items do. Exceptions are distinguished from rule-following items because the exceptions' category assignment runs counter to the rule. According to exemplar models, this reversal is not germane to recognition.

In contrast, a rule-based model, such as RULEX (Nosofsky, Palmeri, & McKinley, 1994), can predict the memory advantage for exceptions. RULEX constructs rules and stores exceptions to the rules. Rule-following items are not individually stored but rather are captured by the rule. Information about exceptions is explicitly stored. The likelihood of recognizing an item is determined by the items in the exception store. The separate storage of exception information allows RULEX to predict the memory advantage of exceptions.

However, RULEX cannot predict the entire pattern of Palmeri and Nosofsky's results. In addition to the memory advantage for exceptions, they found better recognition for studied rule-following over novel items. RULEX cannot predict the memory advantage for rule-following over novel items, which the context model can predict, because rules encode very little information about rule-following items. Thus, Palmeri and Nosofsky combined RULEX and the context model (see Erikson & Kruschke, 1998 for a similar approach involving knowledge gating). The combined model was able to predict the entire pattern of their recognition data.

The inability of RULEX to predict the memory advantage of rule-following over novel items suggests that humans encode more than just rules to represent rule-following items. In support of this idea, Allen and Brooks (1991) found that even when a rule is explicitly applied to a novel item, humans are still somewhat sensitive to the similarity between the novel item and previously encountered examples. This line of work argues that humans have both rule and exemplar systems for learning and recognition.

An alternative approach is clustering as in SUSTAIN (Love, Medin, & Gureckis, 2004). SUSTAIN has aspects of both rule violation and exemplar memory and can predict better memory for exceptions than studied rule-following items as well as better memory for studied rule-following items than novel items. We predict that introducing exemplar-specific attention to exemplar models will allow those models to be sensitive to rule violation.

## Exemplar-Specific Attention

An exemplar model, such as ALCOVE (Kruschke, 1992), treats exceptions and rule-following items in the same manner because all exemplars share the same attention along a dimension. Consequently, ALCOVE cannot predict the memory advantage of exceptions. In contrast, ALCOVE with exemplar-specific attention (ES-ALCOVE) should be able to predict the memory advantage for exceptions. ES-ALCOVE shifts attention to the rule dimension. To classify an exception in the correct category, ES-ALCOVE will distribute attention to the non-rule dimensions so that the exception is distinguished from the rule-following items from the opposing category. While attention will be distributed to the non-rule dimensions for the exceptions, the rule-following items will receive attention on the rule dimension. ES-ALCOVE could predict a memory advantage for exceptions because the exceptions will be "differentiated" from the rule-following items. We simulate ES-ALCOVE to Palmeri and Nosofsky's results to test this intuition. In the next section, we formalize ES-ALCOVE.

### Formalism

ES-ALCOVE stores every training item as a separate trace. The probability that a stimulus $x_i$ is classified into category A is determined by:

$$P(A|x_i) = \frac{exp[\phi \cdot O_A(x_i)]}{exp[\phi \cdot O_A(x_i)] + exp[\phi \cdot O_B(x_i)]} \quad (1)$$

where the parameter $\phi$ controls the decisiveness of classification response. $O_A(x_i)$ is the category A output activation given a stimulus $x_i$ defined as:

$$O_A(x_i) = \sum_{j=1} \omega_{Aj} \cdot S(x_i, y_j) \quad (2)$$

where $\omega_{Aj}$ is the strength of association between category A and exemplar $j$. $S(x_i, y_j)$ is the similarity between a stimulus $x_i$ and a stored exemplar $y_j$ given by:

$$S(x_i, y_j) = exp[-c \cdot D(x_i, y_j)] \quad (3)$$

where the free parameter $c$ scales the strength of overall similarity. $D(x_i, y_j)$ indicates the distance between $x_i$ and $y_j$ defined by:

$$D(x_i, y_j) = \sum_{k=1} \alpha_{jk} \cdot |x_{ik} - y_{jk}| \quad (4)$$

Table 1: Recognition ratings from Palmeri and Nosofsky (1995) and predicted by ES-ALCOVE. The fits of ALCOVE are also included for a comparison. Item types are exceptions (Exc), studied rule-following items (Rul), and novel items (Nov).

| | | ES | |
| Item | Observed | ALCOVE | ALCOVE |
| --- | --- | --- | --- |
| Exc | 6.92 | 6.58 | 6.39 |
| Rul | 5.74 | 5.94 | 6.39 |
| Nov | 5.30 | 5.20 | 5.13 |

where $k$ is the number of dimensions and $\alpha_{jk}$ is the attentional weight for the $k$th dimension of an exemplar $y_j$. Unlike dimension-wide attention, in which all stored exemplars share the same attention along a dimension, each stored exemplar selects which of its dimensions will receive attention.

The familiarity $F(x_i)$ of a stimulus $x_i$ is determined by the summed similarity of the stimulus item to the stored exemplars of both categories A and B:

$$F(x_i) = \sum_{j=1} S(x_i, y_j) \qquad (5)$$

where $S(x_i, y_j)$ is defined in Equation 3.

## Simulation 1

ES-ALCOVE was fit to the mean recognition ratings provided by human subjects in Palmeri and Nosofsky (1995). Prior to recognition, the models learned to classify the members of Categories A and B. Most of the members followed the imperfect rule, but each category contained an exception that violated the rule. The models updated the association ($\omega$) and attention weights ($\alpha$) after the presentation of each training item by gradient descent (see Kruschke, 1992 for derivations) that reduced the sum squared differences between the target and the predicted output values (i.e., Equation 2).

After training, ES-ALCOVE generated recognition ratings. A linear relationship was assumed between the human recognition ratings and ES-ALCOVE's familiarity (i.e., Equation 5). The basic finding was that exceptions received the highest recognition ratings, followed by rule-following items, followed by novel items.

### Results

As shown in Table 1, ES-ALCOVE was able to capture the observed pattern. For a comparison, Table 1 displays that ALCOVE cannot predict the observed results by using the same attention for exemplars representing exceptions and rule-following items. Table 2 shows that as predicted, ES-ALCOVE distributed more attention to the non-rule dimensions

Table 2: The attention weights obtained by ES-ALCOVE for the rule ($D_{rule}$) and the non-rule dimensions (sum for $D_{non-rule}$) of exemplars encoding exceptions (Exc) and rule-following items (Rul).

| Exemplar | $D_{rule}$ | $D_{non-rule}$ |
| --- | --- | --- |
| Exc | .07 | .80 |
| Rul | .22 | .64 |

of the exemplars encoding exceptions than those of the exemplars encoding rule-following items. The opposite pattern was observed for the rule dimension.

The differential attention for exceptions made the exceptions distinctive in memory. Attention scales the distance when the dimension values mismatch. The greater attention for the non-rule dimensions of exemplars encoding exceptions suggests that items with mismatching values on the non-rule dimensions become highly dissimilar. As in RULEX, exceptions are distinguished from rule-following items and remembered better. Unlike RULEX, ES-ALCOVE also correctly predicts a memory advantage for rule-following items over novel items because it retains information about non-rule dimensions.

## Further Test of Exception Memory

ES-ALCOVE was able to predict a memory advantage for exceptions by differentiating the exceptions from rule-following items in terms of attention allocated to exemplars encoding those items. A related finding from the category learning research that examines memory for exceptions is that the memory advantage for exceptions is greater when the violated rule is stronger (Sakamoto & Love, in press).

In Sakamoto and Love, as in Palmeri and Nosofsky, most of the members of categories A and B followed an imperfect rule, and each category contained an exception. Rule strength was manipulated by varying the frequency of rule-following items. Category A contained eight rule-following items, whereas category B contained only four. The classification learning procedure encouraged subjects to entertain the rules "If value 1 on the first dimension, then category A" and "If value 2 on the first dimension, then category B." Category B's exception violated category A's rule, whereas category A's exception violated category B's rule. After training, these exceptions were remembered better than the rule-following items, replicating Palmeri and Nosofsky. Furthermore, as in the schema research (e.g., Rojahn & Pettigrew, 1992), memory for the category B exception, which violated more frequent category A's rule, was enhanced.

As in Palmeri and Nosofsky, exemplar models, such as ALCOVE and the context model, were unable to account for the enhanced recognition of the

category B exception because they did not provide a role for knowledge structures in encoding exceptions and rule-following items. RULEX captured rule-governed behavior with actual rules and was insensitive to the rule frequency manipulation (cf., Pinker, 1991; Smith, Langston, & Nisbett, 1992). Because the context model and RULEX had trouble with different aspects, the combined model was also unable to predict the results.

Although ES-ALCOVE was able to account for the memory advantage for exceptions, it was unable to predict the greater memory advantage for the category B exception. This failure arises because the attention shift in ES-ALCOVE does not distinguish the two exceptions. Because there are more category A rule-following items, it is harder to learn the category B exception than the category A exception. The category B exception results in larger discrepancies between target and predicted output values. ES-ALCOVE treats large and small discrepancies in the same manner and cannot differentiate the two exceptions. As a result, ES-ALCOVE is unable to predict the rule frequency effect observed in Sakamoto and Love.

We created another version of ALCOVE called ESSW-ALCOVE for Exemplar-Specific Squeaky Wheel ALCOVE. In addition to the exemplar-specific attention, ESSW-ALCOVE has a mechanism that emphasizes larger errors and minimizes the impact of smaller ones. ESSW-ALCOVE should be able to predict Sakamoto and Love's results by distributing more attention to the non-rule dimensions of the category B exception, which results in larger errors, than to the non-rule dimensions of the category A exception. ESSW-ALCOVE was fit to Sakamoto and Love's results to test this intuition.

## Simulation 2

In ESSW-ALCOVE, the sum squared error between target and predicted output values minimized by ALCOVE (and ES-ALCOVE) was exaggerated to the 10th power. This causes larger errors to remain relatively large but smaller errors to become extremely small. As a result, relative to ES-ALCOVE, attention learning is accentuated when an item leads to a larger error, and the exemplar representing the error-prone item will receive more uniform attention (hence squeaky wheel). ESSW-ALCOVE can predict Palmeri and Nosofsky's results.

ESSW-ALCOVE learned to classify the training items. Category A contained eight rule-following items, whereas category B contained only four. Each category contained one exception item. After training, ESSW-ALCOVE made two-alternative forced choice judgments. The pairs matched on the rule dimension. The basic finding was that the category B exception, which violated more rule-following items, was remembered better than the category A exception. The probability of choosing the studied

Table 3: Recognition performance (with 95% confidence intervals) observed in Sakamoto and Love (in press) and predicted by ESSW-ALCOVE. The fits of ES-ALCOVE are also included for a comparison. The category B exception (Exc B) violated more rule-following items from category A (Rul A). The category A exception (Exc A) violated fewer rule-following items from category B (Rul B).

| Item | Observed | ESSW ALCOVE | ES ALCOVE |
|------|----------|-------------|-----------|
| Exc B | 87±3 | 85 | 82 |
| Exc A | 79±3 | 80 | 82 |
| Rul B | 69±3 | 70 | 70 |
| Rul A | 70±3 | 70 | 70 |

Table 4: The rank order of ESSW-ALCOVE's attention (with attention weights) for the rule ($D_{rule}$) and the non-rule dimensions (sum for $D_{non-rule}$) of the exemplars encoding the category B exception (Exc B), the category A exception (Exc A), the category B rule-following items (Rul B), and the category A rule-following items (Rul A). The category B exception violated more rule-following items from category A, and the category A exception violated fewer rule-following items from category B. Smaller number represents greater attention.

| Exemplar | $D_{rule}$ | $D_{non-rule}$ |
|----------|------------|----------------|
| Exc B | 4 (0.460) | 1 (0.331) |
| Exc A | 3 (0.463) | 2 (0.314) |
| Rul B | 1 (0.470) | 3 (0.279) |
| Rul A | 2 (0.467) | 4 (0.260) |

(old) item was determined by the exponential decision function:

$$P(old) = \frac{exp[\rho \cdot F(old)]}{exp[\rho \cdot F(old)] + exp[\rho \cdot F(new)]} \quad (6)$$

where $F(old)$ is the model's familiarity (see Equation 5) for the studied item, $F(new)$ is the models' familiarity for the novel item, and $\rho$ is the recognition decision parameter.

### Results

As shown in Table 3, ESSW-ALCOVE, which accentuated errors, captured the observed pattern. For a comparison, Table 3 shows that ES-ALCOVE cannot predict the observed results despite its exemplar-specific attention. The rank order (1 is most and 4 is least) of ESSW-ALCOVE's attention shown in Table 4 reveals that as in Palmeri and Nosofsky, exemplars encoding exceptions resulted in more attention

distributed to the non-rule dimensions than exemplars representing rule-following items. Moreover, as predicted more attention was distributed to the non-rule dimensions for the exemplar encoding category B's exception than for the exemplar encoding category A's exception.

There were more errors involving the category B exception than the category A exception because more rule-following items in category A were similar to the category B exception. Such errors were accentuated in ESSW-ALCOVE, and the category B exception resulted in more uniform attention. ESSW-ALCOVE predicts a memory advantage for the category B exception because it is differentiated from many similar rule-following items from category A.

## Filtration and Condensation

Incorporating exemplar-specific attention and accentuated errors allowed ALCOVE to predict previously challenging data. It is crucial that a model that incorporates new mechanisms can still account for other basic psychological phenomena. Examining filtration and condensation tasks (Gottwald & Garner, 1975; Kruschke, 1993, Matsuka, in press) is important because these tasks investigate how humans allocate attention.

Humans (and ALCOVE) find it easier to learn filtration tasks, in which information from only one dimension is required for perfect classification, than condensation tasks, in which information from two (or more) dimensions is needed (e.g., Kruschke, 1993). This filtration advantage was predicted by ESSW-ALCOVE (and ES-ALCOVE). In filtration, all exemplars attend to the predictive dimension. This leads to increased psychological distances between category A and category B members. In condensation, some items are closer to the opposing category's exemplars even with exemplar-specific attention, and ESSW-ALCOVE has less "confirmations" from some exemplars. ESSW-ALCOVE finds it easier to learn the filtration task, in which it receives more "evidence" for correct category membership.

## Discussion

Humans can flexibly attend to different dimensions of an item depending on the values of the dimensions that are not critical for classification (Aha & Goldstone, 1992). Such flexible attention allowed ES-ALCOVE to differentiate exceptions from rule-following items and predict a memory advantage for exceptions. In ES-ALCOVE, each exemplar selected which dimensions to attend to. ES-ALCOVE attended to the non-rule dimensions of exemplars encoding exceptions but to the rule dimension of exemplars encoding rule-following items. This differential attention made exceptions distinctive in memory.

However, ES-ALCOVE was unable to account for the finding that memory for a violating item is stronger when the violated structure is stronger. This finding was predicted by ESSW-ALCOVE, which accentuated errors. With errors raised to the 10th power, ESSW-ALCOVE distinguished important errors (e.g., miss-classification) from trivial ones (e.g., correct classification with 90% confidence level). ESSW-ALCOVE learned attention more rapidly in response to larger errors and minimized the impact of smaller errors. A similar effect can be obtained by updating the attention weights multiple times on each training trial (e.g., Kruschke, 2001). ESSW-ALCOVE better remembered items that violated a stronger rule because those items were associated with "important" errors.

In addition to the exception memory findings reviewed in this paper, ESSW-ALCOVE was able to predict a filtration advantage. We should note that ALCOVE that accentuated errors without exemplar-specific attention was unable to predict the exception memory findings. Thus, ALCOVE needs both exemplar-specific attention and accentuated errors to account for all of the findings described in this paper. One question to ask is whether these mechanisms are also required by other models.

### Clustering Approach

SUSTAIN (Love et al., 2004), which uses dimension-wide attention, can predict the memory advantage for exceptions as well as the greater memory advantage for exceptions that violate a stronger rule. SUSTAIN clusters together similar items and recruits a new cluster in response to a prediction error. SUSTAIN develops rule-following clusters and shifts attention to the rule dimension. All clusters share the same attention along a dimension. When an exception item elicits a prediction error, SUSTAIN recruits an additional cluster to encode the item. While rule-following items tend to cluster with one another, each exception item will be isolated in its own cluster. This differential storage makes exceptions more distinctive in memory.

SUSTAIN also predicts better recognition for the category B exception that violates more rule-following items from category A. A prediction error occurs when SUSTAIN attempts to cluster together highly similar items from competing categories. The exception clusters brought about such errors by attracting rule-following items from the opposing category. Because there were more category A rule-following items, there were more opportunities for such errors involving the category B exception to occur, and a greater number of category A rule-following clusters were recruited. These clusters formed a highly contrastive backdrop for the category B exception and enhanced recognition.

As ESSW-ALCOVE, SUSTAIN treats "important" and "negligible" errors differently. Large discrepancies between target and predicted output values result in prediction errors. SUSTAIN recruits

new clusters in response to such errors but "ignores" small discrepancies.

SUSTAIN's success in predicting these results using dimension-wide attention suggests that attention mechanisms interact with internal representations of a model (cf., Matsuka, in press). Clustering allows SUSTAIN to be sensitive to rule violation without exemplar-specific attention. In contrast, by storing every studied item, ALCOVE needs exemplar-specific attention to capture the rule-violating nature of exceptions.

## Future Direction

A more flexible model may use dimension-wide or exemplar-specific attention depending on a given task. For example, humans may initially be biased to attend to the same dimensions for all exemplars, analogous to a prior, but over time optimize learning by utilizing a separate attention for certain items. In learning about the rule-plus-exception category structures, humans may initially attend to the rule dimension for all the items. When an exception appears and a prediction error occurs, a separate attention may be used for the exception. Such processes suggest that humans have simplicity bias (cf., Matsuka, 2004). For example, filtration tasks may result in dimension-wide attention, whereas condensation tasks may lead to exemplar-specific attention. Future work should examine when exemplar-specific or dimension-wide attention is more appropriate.

## Acknowledgments

## References

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 534–539). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120*, 3–19.

Barsalou, L. W., & Medin, D. L. (1986). Concepts: Static dimensions or context-dependent representations? *Cahiers de Psychologie Cognitive, 6*, 187–202.

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

Gottwald, R. L., & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception & Psychophysics, 2*, 50–55.

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

Kruschke, J. K. (1993). Three principles for models of category learning. In G. V. Nakamura, R. Taraban, & D. L. Medin (Eds.) *Categorization by human and machines: The psychology of learning and motivation* (Vol. 29, pp. 57–90). San Diego, CA: Academic Press.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45*, 812–863.

Lewandowsky, S., Kalish, M., & Ngang, S. K. (2002). Simplified learning in complex situations: Knowledge partitioning in function learning. *Journal of Experimental Psychology: General, 131*, 163–193.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Human Category Learning. *Psychological Review*. In press.

Matsuka, T. (in press). Generalized exploratory model of category learning. International Journal of Computational Intelligence.

Matsuka, T. (2004). Biased stochastic learning in computational model of category learning. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*.

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., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101(1)*, 53–79.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 548–568.

Pinker, S. (1991). Rules of language. *Science, 253*, 530–535.

Rojahn, K., & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology, 31(2)*, 81–109.

Sakamoto, Y., & Love, B. C. (in press). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*.

Smith, E. E., & Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science, 16*, 1–40.