Journal of Experimental Psychology: General
2004, Vol. 133, No. 4, 534–553

0096-3445/04/$12.00 DOI: 10.1037/0096-3445.133.4.534

Schematic Influences on Category Learning and Recognition Memory

Yasuaki Sakamoto and Bradley C. Love
The University of Texas at Austin

The results from 3 category learning experiments suggest that items are better remembered when they violate a salient knowledge structure such as a rule. The more salient the knowledge structure, the stronger the memory for deviant items. The effect of learning errors on subsequent recognition appears to be mediated through the imposed knowledge structure. The recognition advantage for deviant items extends to unsupervised learning situations. Exemplar-based and hypothesis-testing models cannot account for these results. The authors propose a clustering account in which deviant items are better remembered because they are differentiated from clusters that capture regularities. The function of clusters is akin to that of schemas. Their results and analyses expose connections among research in category learning, schemas, stereotypes, and analogy.

People's ability to categorize underlies many of their cognitive abilities. Classifying a person as a friend, an animal as a dog, and a piece of music as classical are all acts of categorization. For categorization to take place, categories must first be acquired. Consequently, there has been a great deal of interest in understanding how individuals acquire categories from examples. Acquiring new categories necessarily involves changes in memory. The work presented here asks what is stored in memory as a result of category learning. More specifically, the current work explores the effect of category structure and category learning errors on recognition memory.

The acquisition of new categories is often accompanied by category learning errors. For example, a grade school student may categorize a dolphin as a fish. During learning, some category members result in more errors than others. For example, dolphins, whales, and bats are likely to lead to numerous errors when learning about biological categories, whereas sharks, dogs, and robins are less likely to lead to errors. One question we consider is how error rate in acquisition influences subsequent recognition memory.

Determining the relationship between error and recognition rate is not trivial. Other factors are commonly confounded with error rate. For instance, high-error items usually violate a known regularity or rule. A young child may classify all animals that fly as birds. Robins, eagles, and blue jays satisfy this rule, whereas penguins, ostriches, and bats violate this rule and will result in errors upon initial exposure. If a high-error item is better remembered, it is unclear whether this advantage arises from a preponderance of acquisition errors or because the item deviates from an existing knowledge structure (in this case a rule). Furthermore, if violating a known regularity leads to improved memory, how does the nature of the regularity affect memory for deviant items? For instance, the degree of coherence of regularity may play a large role in determining how well violating items are remembered.

Addressing these questions requires an integration of classic work in memory, schema application, stereotypes, and category learning (including computational modeling). We briefly review a subset of relevant work in these areas. Though conceptually related, work in these areas is not as theoretically integrated as one might expect. After consideration of past results, we argue that items are better recognized to the extent that they deviate from an existing and coherent knowledge structure such as a rule, schema, or prototype. Three experiments test our account. Experiment 1's results suggest that the degree of coherence of a knowledge structure influences the recognition of deviating items, which conflicts with a strict rule-based account of knowledge representations. Experiment 2's results demonstrate that a deviant item is better recognized even in the absence of training errors. Finally, Experiment 3's design attempts to separate error rate from violation of a salient knowledge structure. The results suggest that a preponderance of errors can have a role in enhancing recognition memory, but the effect is mediated by subjects' allocation of study time and imposition of structure on the domain.

Our results prove problematic for many existing models of category learning and subsequent recognition. We propose a cluster-based account of category learning that groups related items into a cluster and stresses the role of cluster coherence and differentiation in memory. This account is in accord with our review of research in memory, schemas, and stereotypes.

Background and Motivation

We advance that existing knowledge structures play a role in determining which items are best remembered after a category learning episode. One such candidate structure is a schema. A schema is a general knowledge structure that provides a set of expectations based on prior experience (Brewer & Treyens, 1981;

Yasuaki Sakamoto and Bradley C. Love, Department of Psychology, The University of Texas at Austin.

This work was supported by Air Force Office of Scientific Research Grant FA9550-04-1-0226 and National Science Foundation CAREER Grant 0349101 to Bradley C. Love. We thank Todd Gureckis, Levi Larkey, Todd Maddox, Art Markman, and Thomas Palmeri for their helpful comments.

For additional information, visit http://love.psy.utexas.edu.

Correspondence concerning this article should be addressed to Yasuaki Sakamoto or Bradley C. Love, Department of Psychology, The University of Texas at Austin, 1 University Station A8000, Austin, TX 78712. E-mail: yasu@psy.utexas.edu or love@psy.utexas.edu

Graesser & Nakamura, 1982; Hastie, 1981; Taylor & Crocker, 1981). For example, a person may have a schema for birds that when activated makes properties like flying and laying eggs available. Schemas can guide the encoding and retrieval of information (Alba & Hasher, 1983; Brewer & Nakamura, 1984; Loftus & Mackworth, 1978; Pichert & Anderson, 1977; Srull, 1981). Category information learned from examples can serve similar functions (Goldstone, 1994; Schyns & Murphy, 1994; Wisniewski & Medin, 1994).

Work in the schema and categorization literatures addresses a related set of issues. We draw parallels between these two literatures and argue that findings from both literatures suggest that items tend to be better remembered to the extent that they conflict with an established knowledge structure. It might seem odd that such parallels are not already firmly established. One explanation for the disconnect is the varying methodologies and priorities of the two fields. Work in schemas and stereotypes tends to utilize concepts that are already meaningful to subjects. In contrast, the majority of work in category learning tends to use artificial categories composed of geometric stimuli that have no meaning outside of the experimental context. In a typical category learning experiment, subjects learn to assign geometric stimuli to one of two mutually exclusive categories (e.g., Categories A and B) through trial-by-trial classification learning with corrective feedback (e.g., Estes, 1994; Maddox & Ashby, 1993; Medin & Schaffer, 1978; Nosofsky, 1988). The category learning work that does involve meaningful prior knowledge tends to focus on how such knowledge can facilitate the acquisition of novel categories (Murphy & Allopenna, 1994; Pazzani, 1991; Wattenmaker, Dewey, Murphy, & Medin, 1986). Accordingly, error rate is the primary dependent measure for the majority of work in category learning, whereas measures of recognition and recall figure more prominently in the schema literature. Nevertheless, work from both areas bears on the research questions considered here.

Consistent and Inconsistent Information

One central issue in schema research is whether schema-consistent or schema-inconsistent information is better remembered. For example, encountering a book in a library would be schema consistent (i.e., in accord with expectations), whereas encountering a concert stage would be schema inconsistent. Work in social beliefs and stereotypes has found a memory advantage for schema-consistent information relative to schema-inconsistent information (Rothbart, Evans, & Fulero, 1979; Snyder & Uranowitz, 1978). Social schemas are proposed to function as filtering devices for inconsistent information that lead to inconsistent information being ignored or discounted during the encoding process (Taylor & Crocker, 1981). For example, an accountant's rowdy behavior at a party can simply be ignored or explained away by inferring the accountant was drunk.

The schema-consistent memory advantage has been challenged by other studies that demonstrate that schema-inconsistent information is remembered better than schema-consistent information (Bower, Black, & Turner, 1979; Goodman, 1980; Hastie & Kumar, 1979; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). For example, Hastie and Kumar presented subjects with a list of synonymous adjectives that created a coherent impression of a character. After acquiring this "person schema," behaviors that were inconsistent with this schema were better remembered than those that were consistent.

Rojahn and Pettigrew (1992) conducted a meta-analysis for memory for schema-consistent and schema-inconsistent information and resolved the apparent contradictions across studies. When measures of recognition are corrected for false-alarm rate, schema-inconsistent information is remembered better than schema-consistent information. For a library schema, a common false alarm might be reporting to have seen a book when in fact a book did not appear in any studied scene. Stangor and McMillan (1992) conducted a similar meta-analysis in stereotype research and reached the same conclusion as Rojahn and Pettigrew. This tendency to false alarm to consistent information can also be seen in the Deese–Roediger–McDermott false-memory paradigm (Deese, 1959; Roediger & McDermott, 1995). After correcting for false alarms, the schema-inconsistent memory advantage holds for children, adults, and older adults (List, 1986).

The schema-inconsistent memory advantage may be a specific case of a general advantage for distinctive information. Isolated or deviant events, such as a single word in uppercase in a list of lowercase words, tend to show a recall advantage (Koffka, 1935; von Restorff, 1933; Wallace, 1965). This phenomenon is commonly referred to as the "von Restorff or isolation effect." Unlike typical work in schemas, but like typical work in category learning, subjects gain an appreciation for the structure of the study items during these studies. Once subjects acquire an expectation for the items, the deviant item is analogous to schema-inconsistent information. The von Restorff effect can be seen as a bridge between work in schema research that relies on preexisting knowledge structures and work in category learning in which expectations are developed only after a number of learning trials.

Modulating the advantage of inconsistent information. Experiments in the tradition of von Restorff (1933) and schema research indicate a memory advantage for items that deviate from a salient regularity. One question to consider is whether the strength of the regularity modulates the advantage for deviant items. The strength of the regularity can be manipulated by varying the proportion of items that conform to the regularity. Koffka (1935) reported that when there were more anomalous items in a list, the memory advantage for those items was smaller. Similarly, Rojahn and Pettigrew's (1992) meta-analysis suggests that the memory advantage for the schema-inconsistent items becomes weaker as the proportion of the schema-inconsistent items becomes larger, though the effect was not universal. For example, Pezdek et al. (1989) found that the proportion of inconsistent items had no effect on memory for inconsistent items. One possible explanation for null effects in schema research is that schemas are well learned prior to the experiment and therefore may not be as sensitive to the frequency manipulations experienced in brief laboratory studies.

Processing markers. Inconsistent items may be treated differently from consistent items because inconsistent items are more difficult to process than consistent items (cf., Fabiani & Donchin, 1995; Graesser, 1981). In support of this notion, inconsistent items tend to receive more study time (Stern, Marrs, Millar, & Cole, 1984). This processing account suggests that inconsistent items will not be remembered better than consistent items when study time is limited. In general, people spend more time on difficult study items than on easily mastered items, but this pattern can reverse when study time is limited (Metcalfe, 2002; Thiede &

Dunlosky, 1999). Though deeper processing (cf., Craik & Tulving, 1975) does not necessarily imply increased study time, it seems plausible that if more complete or deeper processing underlies the schema-inconsistent advantage under self-paced study conditions, then limited study time conditions should reduce the memory advantage for schema-inconsistent information. Indeed, people under cognitive load are unable to fully encode schema-inconsistent information and schema-consistent information is better remembered (Brewer & Treyens, 1981; Rojahn & Pettigrew, 1992; Sherman & Frost, 2000). However, when conditions allow for thorough encoding, the memory advantage for schema-inconsistent information is observed.

Related Work in Category Learning

Although the literature on the topic is not as extensive, findings from the category learning research parallel those from the memory research in regards to processing consistent and inconsistent information. In a category learning study involving prior knowledge, Heit (1998) found that subjects weighted inconsistent information more than consistent information when making probability judgments. Furthermore, Heit found that the advantage of inconsistent information was eliminated under speeded study conditions. Unfortunately, most work in the category learning literature is difficult to directly relate to work in the memory literatures as the majority of category learning studies have focused on the difficulty (measured by error rate) of acquiring different category structures (i.e., assignments of stimuli to artificial categories) in the absence of prior knowledge. Thus, the vast majority of studies are not relevant to the current discussion (e.g., Shepard, Hovland, & Jenkins, 1961).

However, research in category learning with prior knowledge can be related to work in schemas. Prior knowledge (somewhat akin to a schema) can assist category learning by making it easier to integrate information about category members (Murphy & Allopenna, 1994). For example, if all the members of Category A have features consistent with a library schema (e.g., books, silent, and desks) and all the members of Category B have features consistent with a concert schema (e.g., bands, alcohol, and loud music), then learning to discriminate between Category A and Category B members would be easier than if the features that predicted A and B were unrelated. Likewise, when expectations about category structures are realized, learning is more efficient (Pazzani, 1991; Wattenmaker et al., 1986).

Unfortunately, these and other studies involving prior knowledge do not directly examine the weighting of consistent and inconsistent information and how recognition memory is affected. These category learning studies do suggest that inconsistent items may require greater attention and may therefore be distinct in memory. In accord with this interpretation, Spalding and Murphy (1996, Experiment 5) found that subjects were unlikely to aggregate items from the same category into one group when category items contained inconsistent information. Subjects treated category items as forming a coherent group only when instructions encouraged subjects to discount inconsistent information by stating that there were mistakes in the printing of the stimuli.

Unlike the previously mentioned category learning studies, Palmeri and Nosofsky's (1995) studies bear directly on how consistent and inconsistent information is encoded. In their studies, subjects learned to classify geometric stimuli into two contrasting categories. The majority of items could be classified by a simple rule (e.g., large items are in Category A, whereas small items are in Category B), while two items were inconsistent with the rule (e.g., a large item belonging to Category B). Items that followed the rule are analogous to schema-consistent information, whereas items that violated the rule are analogous to schema-inconsistent information. Following learning, subjects completed a recognition-rating phase consisting of studied items and novel items constructed by forming novel combinations of the studied items' features. The main finding was that recognition was best for the two rule-violating items. This finding parallels the advantage for schema-inconsistent items found in the memory literatures.

The Palmeri and Nosofsky (1995) studies have a number of strengths relative to schema and stereotype studies. One strength is that the recognition advantage for inconsistent items clearly arises from stimulus encoding and not from retrieval strategies. For example, providing a category label would not improve subjects' ability to recognize a stimulus item. This is important because Anderson and Pichert (1978) have demonstrated that retrieval strategies, such as activating a schema by shifting one's perspective at test, can aid retrieval of studied information. Furthermore, the results are highly interpretable because the stimuli were simple geometric figures that could be counterbalanced and had no prior meaning for subjects.

Palmeri and Nosofsky (1995) modeled their data with the context model (Medin & Schaffer, 1978) and the RULEX (rule-plus-exception) model of category learning (Nosofsky, Palmeri, & McKinley, 1994). The context model is an exemplar model that stores every studied item in memory as a separate trace. Items are represented as vectors of features and are probabilistically classified into Category A or Category B, depending on the item's relative similarity to all exemplars belonging to Categories A and B. The likelihood of recognizing a stimulus as a studied item is proportional to the sum of similarity to all exemplars (from both Categories A and B). The context model alone cannot account for Palmeri and Nosofsky's data, as it predicts no recognition advantage for rule-violating items. This failure arises 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 their category assignment runs counter to the rule and, according to the context model's account of recognition, this reversal is not germane to recognition.

Palmeri and Nosofsky (1995) had more success with the RULEX model. RULEX is a hypothesis-testing model of category learning that constructs rules and stores exceptions to the rules. Rule-following items are not individually stored (as they are in the exemplar model), but rather are captured by the rule. Information about inconsistent items is explicitly stored. The likelihood of recognizing a test item is determined by summing the response from RULEX's rule system (because of the design of Palmeri and Nosofsky's studies, the response is uniform across all items) and the response from the items in the exception store. The storage of exception information allows RULEX to predict a memory advantage for rule-inconsistent information. However, RULEX underpredicts the recognition advantage of rule-following studied items relative to novel items because neither class of items is similar to items in the exception store. Therefore, Palmeri and Nosofsky created a combined model that generates recognition responses by summing the responses of RULEX and the context model (which is sensitive to the difference between studied rule-following items and novel items). This combined model did a good job of accounting for their recognition data.

The inability of RULEX alone to account for the difference in recognition between rule-following and novel items suggests that humans store more than a rule to represent rule-following items. Rules encode little information about rule-following items. 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 (Allen & Brooks, 1991). Analogously, Brooks, Norman, and Allen (1991) demonstrate a tendency to rely on familiar instantiations of abstract features in medical diagnosis. One alternative to a strict rule account is that rule-following items are represented by more schemalike structures such as a cluster.

The SUSTAIN model (the supervised and unsupervised stratified adaptive incremental network model; Love & Medin, 1998; Love, Medin, & Gureckis, 2004) of human category learning proposes such cluster representations. SUSTAIN represents categories by one or more clusters. A stimulus is assigned to the cluster it is most similar to (i.e., closest in multidimensional representational space). SUSTAIN starts with one cluster centered on the first stimulus item encountered. New clusters are recruited in response to surprising events, such as a prediction error in supervised learning or encountering a stimulus that is not similar to any existing cluster in unsupervised learning. Clusters compete to characterize a stimulus with only the most activated (i.e., most similar) prevailing. The response of the winning cluster is attenuated by competition with other clusters. Learning rules update the dominant cluster's position and connection weights. Attention is also adjusted so that the most predictive stimulus dimensions (across all clusters) are the most influential in determining cluster activations. Somewhat analogous to the context model and memory models (e.g., Gillund & Shiffrin, 1984; Hintzman, 1986), recognition is modeled by summing the output of all clusters. SUSTAIN's notion of similarity evolves over time according to the clusters it develops and how it allocates attention.

Rulelike behavior in SUSTAIN is modeled through the development of rule-consistent clusters and the shifting of attention to the rule-relevant stimulus dimension. Unlike RULEX, these clusters encode enough information about the rule-irrelevant dimensions to allow SUSTAIN to predict better recognition for rule-following items than for novel items. Like RULEX, SUSTAIN predicts a recognition advantage for rule-violating items. When a rule-violating item elicits a prediction error (a surprising event), SUSTAIN recruits an additional cluster to encode the item. Although rule-following items tend to cluster with one another, each rule-violating item will be isolated in its own cluster. This differential storage (somewhat analogous to RULEX's exception store) makes rule-violating items more distinctive in memory.

RULEX's and SUSTAIN's treatment of rule-following and rule-violating items parallels findings from the schema research about how consistent and inconsistent information is processed. In particular, RULEX and SUSTAIN are in accord with findings that suggest that people process schema-inconsistent information more deeply and at a greater level of detail. For example, Loftus and Mackworth (1978) found that people fixate more often and longer on schema-inconsistent information than on schema-consistent information. Missing features, new features, or physical changes in the schema-consistent items are rarely noticed, whereas these changes in the schema-inconsistent items are almost always noticed (Friedman, 1979; Goodman, 1980; Heider, 1946; Schank & Abelson, 1977; Sentis & Burnstein, 1979).

Both RULEX and SUSTAIN are in accord with interpretations of the von Restorff (1933) effect that attribute the memory advantage of deviant items to differential attention at encoding (e.g., Green, 1956; Jenkins & Postman, 1948). However, RULEX is at odds with more recent work that demonstrates that deviant items are remembered better, even when presented at the beginning of a study list (e.g., Dunlosky, Hunt, & Clark, 2000; Hunt & Lamb, 2001). SUSTAIN can predict an isolation effect for items presented at the beginning of a list because a deviant item can become distinctive by clustering other items subsequent to the presentation of the deviant item.

Points of Agreement and Contention

Following the schema literature (e.g., Rojahn & Pettigrew, 1992), we propose that items are better remembered to the extent that they are differentiated from a salient knowledge structure during learning. Although RULEX predicts better memory for rule-violating items, it does not predict that the storage of deviant items is affected by the strength of the rule or regularity. One way to manipulate the strength or saliency of a rule is frequency. RULEX posits that actual rules underlie rule-governed behavior. Insensitivity to frequency information is a central property of rules (Pinker, 1991; E. E. Smith, Langston, & Nisbett, 1992). RULEX cannot predict better memory for rule-violating items that violate more frequently encountered rules. In contrast to RULEX, we hypothesize that exceptions differentiated from a rule that many items follow should be better remembered than exceptions differentiated from a rule that few items follow.

SUSTAIN's cluster-based representations are in the spirit of the schema literature. SUSTAIN represents an imperfect rule with one or more clusters, with attention primarily shifted to the rule-relevant dimension. SUSTAIN may be able to predict enhanced memory for items that violate stronger (i.e., more frequent) regularities by recruiting more clusters to represent more salient knowledge structures. Thus, deviant items that violate more salient regularities would be differentiated from a number of highly similar clusters, which would confer a recognition advantage. Experiment 1 tests whether the saliency of knowledge structures affects recognition of deviant items as it does in the memory literatures. The ability of the models to account for the data is also evaluated.

One difference between the memory literatures and the category learning literature is the category learning literature's focus on supervised learning. In fact, RULEX's formalization prevents it from being applied to situations in which feedback is absent. Experiment 2 examines whether an item that deviates from a salient knowledge structure is better remembered without feedback associated with it. Following the schema literature and experiments in the tradition of von Restorff (1933), we predict that feedback is not necessary to induce an advantage for deviant items. SUSTAIN is consistent with this prediction.

In supervised learning situations, inconsistent items tend to be remembered best and result in the most errors during learning.

RULEX and SUSTAIN, in accord with the memory literature, suggest that inconsistent items are remembered best because they violate an existing knowledge structure, not because they result in more errors. In RULEX, inconsistent items violate an existing rule and are preferentially stored as exceptions. Similarly, prediction errors in SUSTAIN occur when an item runs counter to the dominant cluster, which leads to a new cluster being recruited to encode the inconsistent item. Experiment 3 explores the role of errors in a learning task in which there are no salient rule structures. We predict, following RULEX, SUSTAIN, and the memory literatures, that the effect of errors on memory will be mediated through knowledge structures. In the case of Experiment 3, we predict that humans will impose structure on the relatively unstructured categories. Another potential mediator of the effect of errors on memory is study time. As reviewed previously, difficult items tend to receive more study time.

All of our predictions for the three category learning studies follow from work in the memory literatures. Many of these predictions overlap with category learning models, such as SUSTAIN and RULEX. In contrast, all of our predictions run counter to exemplar models. Exemplar models enjoy wide spread acceptance in the category learning community (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986). Unlike the schema literature, exemplar models leave no role for knowledge structures to direct encoding. Before describing the three experiments, the results, and the model fits, we consider the fit of the context model, SUSTAIN, and RULEX to Palmeri and Nosofsky's (1995) data.

Table 1
The Abstract Category Structures Used in Palmeri and Nosofsky's (1995) Experiment 3

Learning item	Dimension value	Novel item	Dimension value
→ A1	21111	N1	12221
A2	11122	N2	12212
A3	12211	N3	12122
A4	11221	N4	12111
A5	12112	N5	11222
A6	11212	N6	11211
A7	12121	N7	11121
A8	11111	N8	11112
→ B1	12222	N9	22221
B2	22211	N10	22212
B3	21122	N11	22122
B4	22112	N12	22111
B5	21221	N13	21222
B6	22121	N14	21211
B7	21212	N15	21121
B8	22222	N16	21112

Note. Stimuli consist of five binary-valued dimensions. There is an imperfect rule on the first dimension, which Items A1 and B1 violate (indicated by the arrows). A = Category A; B = Category B; N = novel item.

Model-Based Accounts of Palmeri and Nosofsky (1995)

In this section, the context model, RULEX, and SUSTAIN are fit to the data from Palmeri and Nosofsky's (1995) Experiment 3. These fits are intended to test our intuitions about the models and to establish the modeling methods used throughout this article. We predict that RULEX and SUSTAIN will predict enhanced recognition for rule-violating items, but the context model will not. The formal descriptions of the models are included in the Appendix, as are the procedures used to obtain model fits. The general philosophy of the model fits was to match the procedures applied to human subjects and models as closely as possible. For instance, the same trial randomization procedures and training criteria were used in the original study and model simulations. In the Palmeri and Nosofsky study, only the recognition data from human subjects completing two consecutive error-free blocks of classification learning (prior to the recognition phase) were included in analyses. This same inclusion criterion was applied to model simulations.

Models were fit to the mean recognition ratings provided by human subjects. Recognition ratings were aggregated by item type. As can be seen in Table 1, Items A1 and B1 violate an imperfect rule on the first stimulus dimension. These two items are referred to as *exceptions.* Items A8 and B8 are referred to as *prototypes* because they follow the imperfect rule and display the modal values of their categories. The remaining items appearing in the learning phase are referred to as *rule-following* items. Items N1–N16 are novel items that served as foils in the recognition phase. These items are grouped according to whether they are similar to one of the exceptions. The basic finding from Palmeri and Nosofsky (1995) is that exception items receive the highest recognition ratings, followed by other items seen in the learning phase, followed by novel items. The human data and model fits are shown in Table 2.

SUSTAIN (four parameters, root mean squared deviation [RMSD] = 0.087) correctly predicts the basic findings. RULEX (five parameters, RMSD = 0.376) correctly predicts the advantage of rule-violating items over rule-following items, but fails to predict an advantage for rule-following items over novel items. The context model (parameter invariant, RMSD = 0.449) complements RULEX by accounting for the overall advantage of studied items while failing to account for enhanced recognition for rule-violating items. Thus, the combined model (six parameters, RMSD = 0.144), which pools the outputs of RULEX and the context model, can account for the basic findings. Table 2 details the models' predictions.

RULEX and SUSTAIN can account for the recognition advantage of inconsistent items (i.e., rule-violating items) over consistent items (i.e., rule-following items) because these models posit knowledge structures that can confer an encoding advantage to deviant items. RULEX forms rules and explicitly stores items that violate the rules in a separate store. SUSTAIN clusters rule-following items and stores rule-violating items, which violate rule-following clusters' expectations, in specifically tuned clusters.

The context model is only sensitive to pairwise similarity relations and does not posit knowledge structures that allow for preferential encoding of rule-violating items. For Palmeri and Nosofsky's (1995) design, the context model's predictions do not depend on parameter settings because each studied item shares the same similarity relations with other items. As in Palmeri and Nosofsky's fit, the context model was unable to predict the recognition advantage for the exception items over the rule-following items, although it captured the higher recognition ratings for the studied items than for the novel items. RULEX cannot predict this general advantage for studied items because recognition in RULEX is driven by its exception store and this information does

Table 2
Human Recognition Ratings Observed in Palmeri and Nosofsky's (1995) Experiment 3 and Recognition Ratings Predicted by the Models

Item	Obs	Con	RUL	C + R	SUS
Exc	6.92	6.39	7.10	6.94	6.88
Pro	6.18	6.39	5.67	6.04	6.16
Rul	5.74	6.39	5.49	5.88	5.88
Sim	5.23	5.13	5.67	5.36	5.24
Dis	5.36	5.13	5.49	5.20	5.26

Note. Item types included exceptions (Exc), prototypes (Pro), studied rule-following items (Rul), novel items similar to an exception (Sim), and other novel items (Dis). Obs, Con, RUL, C + R, and SUS stand for observed data, context model, RULEX (rule-plus-exception model), combined model, and SUSTAIN (supervised and unsupervised stratified adaptive incremental network model), respectively.

not discriminate between rule-following and novel items. In contrast, SUSTAIN's clusters retain the necessary information to support the recognition advantage for rule-following items.

Experiment 1

Experiment 1 tests whether recognition is better for items that violate more salient knowledge structures as in the memory literatures (Koffka, 1935; Rojahn & Pettigrew, 1992). Saliency is operationalized as the number of items that follow a rule. The category structures used in Experiment 1 are similar to those used by Palmeri and Nosofsky (1995). The key difference is that the number of items following the Category A and Category B rules is unequal (see Table 3).

Following the memory literatures, we predict that the exception in the smaller category (i.e., the exception that violates the more frequent rule) will be remembered better than the exception in the larger category. The classification learning procedures lead to predicting the effect in this direction. In classification learning, subjects reason from stimulus values to category membership. As can be seen in Table 3, subjects will entertain the rules "If value 1 on the first dimension, then Category A" and "If value 2 on the first dimension, then Category B." The Category B exception violates the Category A rule, whereas the Category A exception violates the Category B rule. Because the Category A rule is more frequent, we predict enhanced memory for the Category B exception.

Current models should have trouble predicting this pattern of findings. RULEX proposes that actual rules underlie rule-governed behavior and thus should be unable to account for rule frequency effects. SUSTAIN may be able to predict better memory for the exception in the smaller category if more clusters are recruited to represent the imperfect rule for the larger category than for the smaller category. In accord with the schema literature, the cluster encoding the exception in the smaller category would be differentiated from a stronger regularity than the cluster encoding the exception in the larger category. The context model should be unable to account for the predicted results because it does not provide a role for knowledge structures in encoding rule-consistent and rule-inconsistent information.

Method

Subjects. Eighty-two University of Texas undergraduates participated for course credit.

Materials. The experiment was run on Pentium III computers operating in DOS. Data were collected using an in-house, real-time data collection system. The monitors had 15-in. CRT color displays and a refresh rate of 16.67 ms.

The learning phase stimuli were geometric figures that varied in the following five binary valued dimensions: size (small or large), color (blue or purple), border (yellow or white), texture (smooth or dotted), and diagonal cross (present or absent). The five dimensions were mapped (randomly assigned for each subject) onto the logical structure shown in Table 3. The assignment of dimension values was also random for each subject. For example, for some subjects the value 2 on the size dimension signified a large figure, for others it signified a small figure. The five stimulus dimensions are all equally salient and independent (as verified by multidimensional scaling of pairwise similarity ratings—see http://love.psy.utexas.edu/stimuli for details and to download the stimuli).

Design and overview. Subjects completed a learning phase consisting of classification learning trials of the items under the heading Learning item in Table 3. Subjects completed 10 blocks of learning trials. A block is the presentation of each learning item in a random order. After the learning phase, subjects completed a filler phase consisting of three arithmetic problems to prevent rehearsal of information from the learning phase. Then, subjects completed a recognition phase consisting of two-alternative, forced-choice (2AFC) recognition judgments involving items from the learning phase and novel stimulus items. Finally, subjects completed a transfer phase in which they classified items presented in the recognition phase without corrective feedback. The transfer phase allows for evaluation of subjects' learning strategies through examination of how they extend their knowledge to novel items.

The variables item type (rule following or exception) and category size (small or large) were factorially combined. The rule-following items (A2–A9 and B2–B5) followed an imperfect category rule (see Table 3 for the imperfect category rule on the first dimension). Following Medin and Smith (1981) and Palmeri and Nosofsky's (1995) Experiment 1, subjects were provided with a hint to attend to the first dimension. There were two exception items (A1 and B1), one from each category. One category contained nine members (the large category), and the other category contained five members (the small category).

The recognition phase involved forced-choice judgments on 50 pairs of stimuli presented in a random order. Each pair consisted of an item from

Table 3
The Abstract Category Structures Used in Experiment 1

Learning item	Dimension value	Novel item	Dimension value
→ A1	21112	N1	11221
A2	12122	N2	12112
A3	11211	N3	12221
A4	12211	N4	12212
A5	11122	N5	12222
A6	12111	N6	21221
A7	11222	N7	22112
A8	11212	N8	22221
A9	12121	N9	22212
→ B1	11121	N10	22222
B2	22122		
B3	21211		
B4	22211		
B5	21122		

Note. A = Category A; B = Category B; N = novel item. There is an imperfect rule on the first dimension, which Items A1 and B1 violate (indicated by the arrows).

the learning phase and a novel item. Twenty items, 10 studied and 10 novel, were used. The 10 studied items were 5 items from Category A (A1–A5) and 5 from Category B (B1–B5). The 10 novel items are displayed under Novel item in Table 3. Because items with value 1 on the first dimension were more frequent than items with value 2 during the learning phase, the false-alarm rate for recognizing the items with value 1 on the first dimension would be higher than that for recognizing the items with value 2 when the items are judged in isolation. To avoid excessive false alarms, items were paired to match on this dimension. The 5 studied items with value 1 on the first dimension (i.e., A2–A5 and B1) were paired with each of the 5 novel items with value 1 on the first dimension (i.e., N1–N5), which resulted in 25 pairs. Another set of 25 pairs was created in the same manner using the items with value 2 on the first dimension.

In the transfer phase, subjects classified the same 20 stimuli used in the recognition phase without corrective feedback. Subjects completed two blocks of transfer trials.

Procedure. Instructions for the learning, filler, recognition, and transfer phases were displayed on the monitor at the start of each phase. The background color was black.

On each trial in the learning phase, one stimulus appeared at the center of the monitor and the text "Category A or B?" was displayed above the stimulus. In addition, a hint to attend to the rule-relevant dimension was presented. For example, "Look whether the size is small or large" appeared above the text "Category A or B?" when size was the rule-relevant dimension. The instruction stated that this strategy may not work all the time. Subjects indicated their category membership judgment by pressing the *A* or *B* key. After subjects responded, the text and the hint above the stimulus were replaced with visual (e.g., "Right! The correct answer is A" or "Wrong! The correct answer is B") and auditory corrective feedback (i.e., a low-pitch tone for errors and a high-pitch tone for correct responses). The stimulus and the visual feedback were displayed for 2,501 ms (150 screen refreshes) after subjects responded. A blank screen was then displayed for 834 ms (50 screen refreshes) and the next trial began.

After completing the learning phase, subjects were presented with a series of three arithmetic problems. Each problem consisted of two integers (randomly generated between 10 and 49) presented side by side (e.g., 22 + 34 = ?); the problem remained displayed until the subjects responded. Subjects received both auditory and visual feedback indicating whether they added the numbers correctly.

In the recognition phase, a pair of stimuli was presented side by side on each trial. Each pair consisted of a learning item and a novel item, as described earlier. The text "Old: left (Q) or right (P)?" was displayed above the stimuli. Subjects pressed the *Q* key (on the left side of the keyboard) to indicate the left item was old (appeared in the learning phase) and pressed the *P* key (on the right side of the keyboard) to indicate that the right item was old. For each pair, the studied and novel items were randomly assigned to the left or to the right position. No corrective feedback was given to subjects. After subjects responded, the text "Thank you" appeared below the stimulus and a high-pitch tone sounded. The stimulus and the text "Thank you" were displayed for 2,501 ms after subjects responded. A blank screen was then displayed for 834 ms and the next trial began.

Experiment 1 featured a transfer phase following the recognition phase. The procedure for the transfer phase was similar to that for the learning phase except that no hint or feedback was provided. After participants responded either "A" or "B," a high-pitch tone sounded and the text "Thank you" was displayed below the stimulus.

Results

One subject did not perform above chance (i.e., 50%) in the learning phase and was excluded from further analysis.[1] Data from the learning, recognition, and transfer phases were analyzed, although the data of primary interest were from the recognition phase. Table 4 displays subjects' mean accuracies in the learning, recognition, and transfer phases.

Table 4
Mean Accuracies in the Learning, Recognition, and Transfer Phases

Item	Learning	Recognition	Transfer
Exc S	.44	.87	.64
Exc L	.46	.79	.56
Rul S	.86	.69	.85
Rul L	.92	.70	.86

Note. Item types included the exception from the small category (Exc S), the exception from the large category (Exc L), the rule-following items from the small category (Rul S), and the rule-following items from the large category (Rul L).

Learning phase. A factorial Category Size (small or large) × Item Type (rule following or exception) analysis of variance (ANOVA) was performed on the accuracy data from the learning phase. Subjects were more accurate (.69 vs. .65) for the large-category items than for the small-category items, $F(1, 80) = 7.37$, $MSE = 0.02$, $p < .01$. The effect size, measured by partial η^2 (for significant ANOVA results for all experiments), was .08, suggesting that the category size by itself accounted for only 8% of the overall (effect plus error) variance. As predicted, subjects were less accurate (.45 vs. .90) on the exceptions than on the rule-following items, $F(1, 80) = 319.49$, $MSE = 0.05$, $p < .001$, with partial $\eta^2 = .80$. The interaction did not reach significance, $F(1, 80) = 1.75$, $MSE = 0.02$, $p \approx .19$. The difference in subjects' accuracy between the small-category exception and the large-category exception (see Table 4) was not significant ($t < 1$). However, the difference in subjects' accuracy between the small-category rule-following items and the large-category rule-following items was significant, $t(80) = 5.82$, $p < .001$.

Unbeknownst to subjects, response time was collected in the learning phase. A factorial Category Size × Item Type ANOVA was performed on response time (collapsed across accurate and inaccurate responses). Though not predicted, subjects spent less time (1,872 ms vs. 1,980 ms) considering items in the small category than items in the large category, $F(1, 80) = 5.43$, $MSE = 172{,}224$, $p < .05$, with partial $\eta^2 = .06$. As predicted, the exceptions resulted in greater response times (2,020 ms vs. 1,824 ms) than the rule-following items, $F(1, 80) = 14.65$, $MSE = 195{,}768$, $p < .001$, with partial $\eta^2 = .15$. The interaction between category size and item type was significant, $F(1, 80) = 14.82$, $MSE = 340{,}595$, $p < .001$, with partial $\eta^2 = .16$. For the exception items, response time was 358 ms faster when belonging to the small category than to the large category (1,841 ms vs. 2,199 ms). In contrast, response time for the rule-following items was 142 ms faster when belonging to the large category than to the small category (1,761 ms vs. 1,903 ms). The interaction indicates that items displaying the more frequent value on the first dimension had faster response times. Subjects spent significantly more time on the exception in the large category than on the exception in the

[1] Including such subjects does not change the pattern of results in any experiment.

small category, $t(80) = 3.48$, $p < .001$. Response time was greater for rule-following items in the small category than for rule-following items in the large category, $t(80) = 3.07$, $p < .01$.

Recognition phase. The recognition results are shown in Figure 1. A factorial Category Size × Item Type ANOVA was performed on 2AFC recognition accuracy data. Subjects were more accurate (.78 vs. .74) with items in the small category than with items in the large category, $F(1, 80) = 5.28$, $MSE = 0.02$, $p < .05$, with partial $\eta^2 = .06$. As predicted, the exceptions were better remembered (.83 vs. .70) than the rule-following items, $F(1, 80) = 39.31$, $MSE = 0.04$, $p < .001$, with partial $\eta^2 = .33$. As predicted, there was a significant Category Size × Item Type interaction, $F(1, 80) = 7.25$, $MSE = 0.02$, $p < .01$, with partial $\eta^2 = .08$. For the exception items, recognition was 8% better for the small category than for the large category; for rule-following items, recognition was 1% worse for the small category than for the large category (see Table 4). In accord with our main prediction, subjects remembered the exception from the small category better than the exception from the large category, $t(80) = 2.72$, $p < .01$. The difference between rule-following items from the small and large categories was not significant ($t < 1$).

Transfer phase. For the purposes of analyses, novel items were considered to be in the category for which they satisfied the imperfect rule. A factorial Category Size × Item Type ANOVA was performed on transfer accuracy data. Subjects were not significantly more accurate (.74 vs. .71) for small-category items than for large-category items, $F(1, 80) = 2.15$, $MSE = 0.04$, $p \approx .15$. As predicted, the exceptions resulted in lower accuracy (.60 vs. .86) than the rule-following items, $F(1, 80) = 28.36$, $MSE = 0.19$, $p < .001$, with partial $\eta^2 = .26$. The interaction between category size and item type approached significance, $F(1, 80) = 3.28$, $MSE = 0.04$, $p \approx .07$. For the exception items, accuracy was 8% better for the small category than for the large category; for rule-following items, accuracy was 1% worse for the small category than for the large category (see Table 4). The difference in subjects' accuracy between the small-category exception and the large-category exception approached significance, $t(80) = 1.75$, $p \approx .08$. There was no significant difference in subjects' accuracy between the small-category and the large-category rule-following items ($t < 1$).

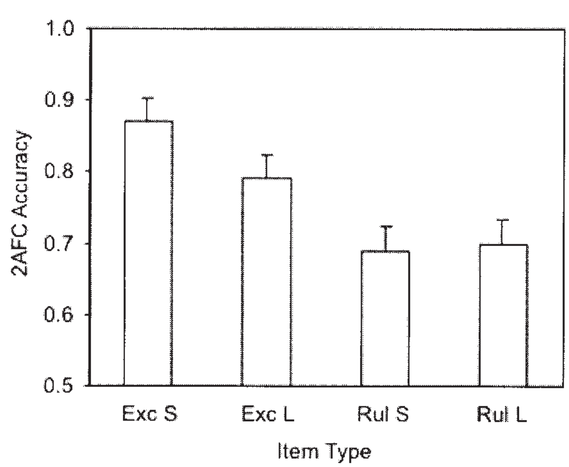

Figure 1. Mean accuracies in the recognition phase of Experiment 1 are shown along with 95% within-subject confidence intervals (see Loftus & Masson, 1994). Error bars represent standard errors of the mean. Exc S = exception of the small category; Exc L = the exception of the large category; Rul S = the rule-following items of the small category; Rul L = the rule-following items of the large category; 2AFC = two-alternative forced choice.

Model Analyses

As predicted, people remembered the exception in the smaller category better than the exception in the larger category. As the fits shown in Table 5 indicate, this pattern of results is problematic for the context model (two parameters, RMSD = 0.033), RULEX (six parameters, RMSD = 0.053), and the combined model (eight parameters, RMSD = 0.033). SUSTAIN (four parameters, RMSD = 0.027) successfully accounts for the pattern of results.

The context model, RULEX, and the combined model all predicted virtually equal recognition for both exceptions. The context model was able to predict better recognition for exceptions overall because of its multiplicative rule for calculating similarity. Although all items had an equal number of matches on average with other items, matches for exceptions were concentrated among fewer items, which is favored by the context model's multiplicative similarity rule. As in the Palmeri and Nosofsky (1995) simulation, RULEX formed a rule on the first dimension and stored the two exceptions.

SUSTAIN captures the pattern of results, and its predictions lie within the error bars for the human data shown in Figure 1. SUSTAIN recruited a cluster for each exception. Of importance, SUSTAIN recruited more clusters to encode the rule-following items from the larger category than it recruited to encode the rule-following items from the smaller category. SUSTAIN tends to cluster together highly similar items from the same category. When SUSTAIN tries to cluster together highly similar items from competing categories, a prediction error occurs (i.e., a surprising event) and a new cluster is recruited. In the simulations of Experiment 1, the exception clusters brought about such errors by attracting rule-following items from the opposing category. Because there were more rule-following items in the larger category, there were more opportunities for such errors involving the exception from the small category to occur. As a result, a greater number of rule-following clusters for the large category were recruited. These clusters formed a highly contrastive backdrop for the exception in the smaller category and allowed SUSTAIN to predict enhanced recognition of that item.

Discussion

The results from Experiment 1 parallel those from the memory research. Exception items (akin to schema-inconsistent information) that were differentiated from a salient knowledge structure were remembered best. The exception in the smaller category that violated a more frequent (i.e., salient) regularity was remembered better than the exception in the larger category. RULEX and the context model could not account for the results, whereas SUSTAIN could by recruiting more clusters to represent rule-following items from the larger category, which created a strong backdrop for the exception in the smaller category to be recognized against. The results of Experiment 1 and the success of

Table 5
Two-Alternative, Forced-Choice Recognition Performances Observed in Experiment 1 (With 95% Confidence Intervals) and Predicted by the Models

Item	Obs	Con	RUL	C + R	SUS
Exc S	.87 ± .03	.83[a]	.86	.83[a]	.90
Exc L	.79 ± .03	.83[a]	.85[a]	.83[a]	.77
Rul S	.69 ± .03	.70	.65[a]	.69	.71
Rul L	.70 ± .03	.70	.65[a]	.69	.67

Note. Item types included the exception from the small category (Exc S), the exception from the large category (Exc L), the rule-following items from the small category (Rul S), and the rule-following items from the large category (Rul L). Obs, Con, RUL, C + R, and SUS stand for observed, context model, RULEX (rule-plus-exception model), combined model, and SUSTAIN (supervised and unsupervised stratified adaptive incremental network model), respectively.
[a] The predicted value falls outside of the confidence intervals.

SUSTAIN suggest that rules might be represented by cluster or schemalike structures, as opposed to actual rules.

Although RULEX failed to account for the recognition advantage of the exception in the smaller category, its simulations were theoretically illuminating. RULEX attempts to store more information about an exception when the exception is more confusable with opposing rule-following items. Thus, RULEX can be sensitive to the diversity of items in the opposing category. In Experiment 1 (see Table 3), rule-following items in the larger category were both more frequent and more diverse in their combination of feature values than rule-following items in the smaller category.

This analysis indicates that there are two senses of "more." Greater frequency corresponds to more rule-following tokens, whereas greater diversity corresponds to more rule-following types (cf. Barsalou, Huttenlocher, & Lamberts, 1998). Like Experiment 1, these two notions of more have perfectly co-occurred in the memory literatures. Although Experiment 1 was motivated by work in the memory literatures, perhaps our analysis can illuminate future schema research. Preliminary results from category learning experiments suggest that when either frequency or diversity is held constant, there is an independent effect of the other on the recognition of deviant items (Love & Sakamoto, 2003).

We have focused on structural explanations based on differentiation to explain the recognition memory advantage of the exception in the smaller category. Other explanations based on study time or error rate are more appropriately considered correlated measures rather than true explanations. Nevertheless, these explanations fall short in Experiment 1. Study time was greater for the exception in the larger category than for the exception in the smaller category. Error rate was also not predictive of the recognition advantage of the exception from the smaller category over the exception from the larger category. However, as in Palmeri and Nosofsky (1995), both exceptions did result in more errors than rule-following items and were remembered better.

In Experiment 2, we examined whether structure violation by itself can lead to memory advantage for deviant items by utilizing an unsupervised category learning paradigm (cf. Ashby, Queller, & Berretty, 1999; Love, 2003b). Clearly, errors are not relevant for explaining the outcome of Experiment 2 as no corrective feedback is provided in unsupervised learning.

Experiment 2

In contrast to Experiment 1, subjects in Experiment 2 learned about members of a single category in an unsupervised fashion and imposed their own organization on the stimulus items (cf. Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987). In the learning phase, subjects viewed members of Category A and rated their typicality. Following the learning phase, subjects' ability to recognize studied stimuli was measured.

Another key difference between Experiments 1 and 2 is that the stimulus dimensions are intercorrelated in Experiment 2. Across dimensions, stimulus value 2 tends to co-occur with value 2 (see Table 6). This rich structure is in accord with the structure of natural categories (Rosch & Mervis, 1975) and the idea that a schema is a set of related expectations.

Adopting this analysis, Items A2–A6 in Table 6 can be viewed as schema consistent, whereas Item A1 can be viewed as schema inconsistent. No single rule defines the schema. The schema-consistent items form a cluster in which the deviant Item A1 is not easily accommodated. Therefore, we predict that the deviant item will be better recognized than the schema-consistent items. Similarly, the improved memory for the deviant item could be understood as a von Restorff (1933) effect in which the isolated item is better remembered. Critically, Item A1 is only isolated as a result of learning to relate the other items to one another in an unsupervised fashion. The novel items (N1–N6 in Table 6), which served as foils in the recognition phase, follow the same frequency distribution as the studied items. Across the novel and studied item stimulus sets, each stimulus dimension displays the 1 value two times and the 2 value four times.

If subjects simply formed a prototype (i.e., an all inclusive schema) during the learning phase, recognition of the deviant item should be worse than the other studied items because the deviant item would not be familiar when compared with the stored prototype. In fact, the deviant item would be less familiar than the novel items. Exemplar-based approaches make similar predictions based on the sum of item similarities. In contrast, SUSTAIN's predictions are aligned with the schema literature. SUSTAIN predicts that schema-consistent Items A2–A6 will form one cluster, whereas deviant Item A1 will be stored in its own cluster. SUSTAIN predicts that isolated Item A1 will be remembered best.

Method

Subjects. Fifty-two University of Texas undergraduates participated for course credit.

Table 6
The Abstract Category Structure Used in Experiment 2

Learning item	Dimension value	Novel item	Dimension value
→ A1	11111	N1	22222
A2	22221	N2	22211
A3	22212	N3	22112
A4	22122	N4	21122
A5	21222	N5	11222
A6	12222	N6	12221

Note. There is a family resemblance structure for the Category A (A) items, which Item A1 violates (indicated by the arrow). N = novel item.

Materials. The apparatus was the same as that used in Experiment 1. The stimuli were drawn from the same set used in Experiment 1.

Design and overview. Subjects completed a learning phase in which they rated the typicality (explained as goodness of example in the instructions) of the learning items shown in Table 6 followed by the same filler phase used in Experiment 1. Finally, in the recognition phase, subjects judged whether all of the stimuli shown in Table 6 appeared in the learning phase (two blocks).

Prior to making typicality ratings in the learning phase, subjects were familiarized with the six learning phase stimuli through their sequential presentation (1 block) in a random order. Following this familiarization, subjects completed 10 blocks of typicality ratings.

Unlike Experiment 1, all measures (i.e., typicality and recognition) involved a 29-point rating scale (displayed horizontally on the monitor). The advantage of rating scales lies in greater statistical power. To minimize artifacts, the polarities of the scales were counterbalanced. For example, for some subjects the right end of the scale indicated a typical stimulus, whereas for others the left end did. Likewise, the right end of the recognition scale indicated *new* for some subjects, whereas the left end did for others. Thus, simply providing the same or opposite responses in the learning and recognition phases would result in a null effect. For the purposes of analyses, all ratings were mapped onto a 0 (*atypical or new*) to 1 (*typical or old*) scale with .5 as the midpoint.

Procedure. At the start the learning phase, there was a familiarization block in which the subjects simply examined the members of Category A. A stimulus was displayed for 5,002 ms. The text ("Members of Category A") above the stimulus notified subjects that all of the items belonged to Category A. A blank screen was then displayed for 834 ms and the next stimulus was presented.

After the familiarization block, subjects rated how typical of an example each member of Category A was using the typicality rating scale described earlier. The scale was displayed beneath the stimulus. Initially, there was a red ball at the center of the scale. The ends of the scale were labeled *Good example* and *Bad example*. Subjects used the less than and greater than keys to move the ball toward the desired rating. Subjects pressed the *Z* key when the ball was in the desired position. After pressing the *Z* key, a high-pitch tone sounded and the text "Thank you" was presented below the stimulus. The stimulus and the text were displayed for another 2,501 ms. A blank screen was then displayed for 834 ms and the next trial began. In the recognition phase of Experiment 2, subjects indicated their recognition rating using the same procedure as was used to elicit typicality ratings, but the ends of the scale were labeled *Old* and *New*.

Results

All subjects were included in the analyses. As predicted, subjects rated structure-following Items A2–A6 as more typical than the deviant Item A1 (.43 vs. .23) in the learning phase, $t(51) = 6.53$, $p < .001$.

The main prediction was that the deviant item would be recognized best. The mean recognition ratings are shown in Figure 2. A significant effect of item type was observed, $F(2, 102) = 42.12$, $MSE = 0.01$, $p < .001$, with partial $\eta^2 = .45$. Planned comparisons revealed that subjects rated the deviant item as older than the structure-following items (see Table 7), $t(51) = 4.41$, $p < .001$. The deviant and the structure-following items were both rated older than the novel items, $t(51) = 7.79$, $p < .001$ and $t(51) = 6.06$, $p < .001$, respectively.

Although our main prediction was confirmed, it is not clear whether the deviant item stood out from a schemalike or rulelike structure. Previously, Medin and his colleagues (Ahn & Medin, 1992; Medin et al., 1987) found in free-sorting tasks that people's sorting behavior is often based on a single dimension. Attending to

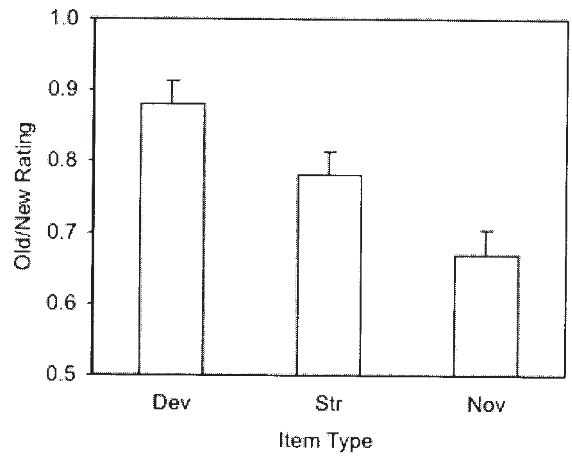

Figure 2. Mean ratings in the recognition phase are shown along with 95% within-subject confidence intervals (Loftus & Masson, 1994). Error bars represent standard errors of the mean. Dev = the deviant item; Str = the structure-following items; Nov = the novel items.

a single dimension in Experiment 2 would lead to one of the structure-following items violating the rule in addition to the deviant item (see Table 6). Each subject's recognition ratings for the five structure-following items were sorted. If subjects formed a single-dimension rule, the difference between the item rated oldest and the item rated second oldest should be greater than the difference between the second and third oldest. The results did not support this prediction. On average, the change from the oldest to the second oldest rated item (.05) was not statistically different from the change from the second to the third oldest rated item (.06; $t < 1$).

Model Analyses

The fits of the context model and SUSTAIN to the recognition phase data are shown in Table 7. Although RULEX and the combined model were not fit because RULEX is not readily applied to an unsupervised learning situation, the preceding analysis inspired by RULEX conflicts with the notion that subjects applied single-dimension rules. SUSTAIN (four parameters, RMSD = 0.002) captured the pattern of data, whereas the context model (one parameter, RMSD = 0.052) failed to predict the recognition advantage of the deviant item over the other studied items.

The context model's one applicable parameter governs its sensitivity to feature mismatches in calculating pairwise similarity. The best fit of the context model displayed extreme sensitivity, which leads to minimal stimulus generalization. Thus, all studied items were recognized at the same rate. Decreasing sensitivity and increasing stimulus generalization would result in the deviant item being remembered worse than other items.

SUSTAIN's modal solution was to create one cluster for the schema-consistent Items A2–A6 and a separate cluster for the deviant Item A1. This solution led to a large recognition advantage for the deviant item. Another popular solution was to aggregate all studied items into one cluster. This solution predicts a recognition disadvantage for the deviant item. Individual differences in

Table 7
Recognition Ratings Observed in Experiment 2 (With 95% Confidence Intervals) and Predicted by the Context Model and SUSTAIN

Item	Obs	Con	SUS
Dev	.88 ± .03	.84[a]	.88
Str	.78 ± .03	.84[a]	.78
Nov	.67 ± .03	.65	.67

Note. Item types included deviant item (Dev), structure-following items (Str), and novel items (Nov). Obs, Con, and SUS stand for observed, context model, and SUSTAIN (supervised and unsupervised stratified adaptive incremental network model), respectively.
[a] The predicted value falls outside of the confidence intervals.

SUSTAIN simulations are in accord with our hypothesis that differentiation of deviant items is necessary for a recognition advantage to occur.

Discussion

The results from Experiment 2 support our basic hypothesis that items are better remembered to the extent that they violate a salient knowledge structure. In the case of Experiment 2, the salient knowledge structure was acquired in the absence of supervision and the violating item was not associated with errors.

Unlike the exception items from Experiment 1, the deviant item did not violate a rule, but instead stood out against a pattern of expectations that the other items largely satisfied. Rather than being integrated with the other items, the deviant item was isolated and better remembered. This interpretation is in accord with SUSTAIN's clustering account. SUSTAIN's typical solution was to devote one cluster to encoding the five items that were consistent with the characteristic pattern (akin to a schema) and another cluster to encoding the deviant item.

The context model posits no knowledge structures that can confer an encoding advantage to deviant items. The context model was unable to predict that the deviant item was remembered better than the structure-following items because the deviant item displayed atypical values on all stimulus dimensions.

The results from Experiments 1 and 2 suggest that errors are not necessary to obtain a recognition advantage for deviant items. In Experiment 1, one exception was recognized better than the other exception even though their error rates were statistically equivalent. Similarly, the deviant item in Experiment 2 was recognized best with no errors associated with any item. Although these findings demonstrate that recognition can vary somewhat independently of errors, error rate was not directly manipulated. In contrast to Experiments 1 and 2's focus on category structure, Experiment 3 investigates whether making an error by itself leads to improved recognition.

Experiment 3

Experiment 3 explores the role of category learning errors on subsequent recognition memory. Many learning models suggest that errors play a central role in learning. For example, the Rescorla and Wagner (1972) model predicts that associations are strengthened to the extent that expectations do not match outcomes. Subsequent error-minimization models follow a similar logic (Gluck & Bower, 1988; Kruschke, 1992; Rumelhart, Hinton, & Williams, 1986). These models are partially motivated by conditioning phenomena, such as the blocking effect (Kamin, 1969), which suggest that errors are necessary for changes in memory. Somewhat related attentional models posit that errors mediate memory storage by leading to greater focus on error-producing items (Mackintosh, 1975; Pearce & Hall, 1980).

Findings from neuroscience are consistent with error-driven models of learning. Norepinephrine neurons show transient activity whenever reinforcement contingencies change during acquisition (Sara & Segal, 1991). Similarly, cerebellar climbing fibers and neurons in dorsolateral prefrontal cortex appear sensitive to an error signal (Ito, 1989; Niki & Watanabe, 1979). On the reward side of the equation, dopamine neurons in the substantia nigra are highly activated in the presence of a surprising reward, but are not active for anticipated rewards (see Schultz, 2000, for a review). Work in event-related potentials provides further support for error-driven learning models (Holroyd & Coles, 2002; Kopp & Wolff, 2000).

Despite all of this work, it is not clear whether errors per se drive learning because errors typically occur when a stimulus runs counter to an existing knowledge structure such as a rule. For example, in Experiment 1 and in Palmeri and Nosofsky's (1995) experiments, exception items violated a salient rule and also resulted in a large number of training errors. Furthermore, the results from Experiment 2 demonstrated that errors are not necessary for structure-violating items to display a recognition advantage.

Experiment 3 complements Experiment 2 by considering the role of errors in recognition with irregular category structures that lack natural clusters, imperfect rules, and so forth. Whereas Experiments 1 and 2 manipulated the structural role of stimulus items (e.g., schema consistent or schema inconsistent), Experiment 3 manipulates the error rate of stimulus items. Error rate was manipulated by creating a stimulus item that always resulted in a learning error (i.e., when the subject responds "A" the feedback is "B," and vice versa). Another stimulus item always resulted in correct feedback. The remaining items received regular feedback. The category membership of the regular feedback items was not contingent on the subject's response. If errors lead to improved recognition memory, the item that always results in an error should be remembered best, followed by regular feedback items (which on average will elicit a moderate number of errors), followed by the item that always results in affirmative feedback.

RULEX and SUSTAIN make more subtle predictions about the role of errors in memory. These models suggest that human learners will impose structure on learning problems even when the category structures are highly irregular. Items that violate these imposed structures will result in errors and will be better remembered. Thus, these models hold that the effect of errors on recognition is mediated through structure violation. SUSTAIN makes an error when the current item runs counter to the dominant cluster, whereas RULEX makes an error when the current item runs counter to the applicable rule. According to SUSTAIN and RULEX, errors cannot be completely dissociated from structural considerations. The context model provides no role for errors and predicts that the always-wrong item, regular feedback items, and the always-right item will be remembered at the same rate.

Method

Subjects. Seventy-nine University of Texas undergraduates participated for course credit.

Materials. The apparatus was the same as that used in Experiment 1. The stimuli were drawn from the same set used in Experiment 1. Four of the five dimensions were randomly selected for each subject, with the remaining dimension fixed to one of its two values (also randomly determined).

Design and overview. Subjects completed a learning phase consisting of classification learning trials involving the eight items under the heading Learning item in Table 8. Subjects completed 20 blocks of learning trials with corrective feedback. Following the learning phase, subjects completed the same filler phase used in Experiment 1. Finally, subjects completed 2 recognition blocks of 2AFC trials. Each block was composed of eight pairs of items. Each of the learning items was paired with a novel item under the heading Novel item in Table 8. The paired items differed on the last dimension value.

The variable of primary interest was the feedback type (regular, always wrong, or always right) of learning phase items. Regular feedback items (A1–A3 and B1–B3 in Table 8) generated feedback in accord with the category assignment shown in Table 8. The always-wrong item always generated negative feedback (i.e., when the subject responded "A" the feedback was "B," and vice versa). In contrast, the always-right item always generated positive feedback (i.e., the feedback was always in accord with the response).

Procedure. The procedure in Experiment 3 was identical to that in Experiment 1 except for the following: (a) no hint was provided during the learning phase and (b) there was no transfer phase.

Results

Seven subjects were excluded from further analysis. One subject did not complete the experiment. The other 6 did not perform above the chance level of 50% for regular feedback items in the learning phase.

Learning phase. Unbeknownst to subjects, response time was collected in the learning phase. The response time (collapsed across accurate and inaccurate responses) was 1,858 ms for the always-wrong item, 1,653 ms for the regular feedback items, and 1,472 ms for the always-right item, yielding a significant effect of item type, $F(2, 142) = 31.83$, $MSE = 2{,}683{,}346$, $p < .001$, with partial $\eta^2 = .31$. In the learning phase, subjects devoted more time to the always-wrong items than to the regular feedback items, $t(71) = 5.23$, $p < .001$, and the always-right items, $t(71) = 6.63$, $p < .001$. Subjects responded more slowly to the regular feedback item than to the always-right item, $t(71) = 3.95$, $p < .001$.

Table 8
The Abstract Category Structures Used in Experiment 3

Learning item	Dimension value	Novel item	Dimension value
A1	1111	N1	1112
A2	2121	N2	2122
A3	1212	N3	1211
B1	2222	N4	2221
B2	1221	N5	1222
B3	2112	N6	2111
→ W	1122	N7	1121
R	2211	N8	2212

Note. Item W always resulted in negative feedback (indicated by the arrow). A = Category A; B = Category B; W = wrong; R = right.

Recognition phase. The mean recognition accuracies are shown in Figure 3. A significant effect of feedback type was observed, $F(2, 142) = 6.09$, $MSE = 0.64$, $p < .01$, with partial $\eta^2 = .08$. Planned comparisons revealed that subjects demonstrated significantly higher recognition accuracy for the always-wrong item than for the regular feedback item (see Table 9), $t(71) = 3.85$, $p < .001$, and the always-right items, $t(71) = 2.39$, $p < .05$. There was no significant recognition difference between the always-right and regular feedback items ($t < 1$).

The poor recognition performance for items other than the always-wrong item is partially attributable to similarity of targets and foils, which only differed on one dimension. Another possible factor is subjects' imposition of structure on Categories A and B, which we consider next.

Imposing structure on learning and recognition. SUSTAIN and RULEX predict that subjects will impose structure on relatively unstructured learning problems and that the effect of errors on recognition memory will be mediated through these imposed structures. In this light, the always-wrong item should be remembered best because it violates any imposed structure. Analyses reported here for the regular feedback items further support the view that subjects are imposing structure on Categories A and B.

Imposing structure in terms of rules or tuned clusters results in a bimodal distribution of errors, with consistent items leading to few errors and inconsistent items leading to many errors. In contrast, the null hypothesis holds that items are encoded independently of one another at the same rate and that responses are distributed accordingly. The last two blocks of the learning phase were analyzed to test whether the pattern of responding for the regular feedback items followed the predictions of the structure imposition view. As predicted by this view, the distribution of item responses for the last two blocks (298 consecutive corrects vs. 277.12 expected, 96 one correct/one error vs. 137.76 expected, and 38 consecutive errors vs. 17.12 expected) was more skewed toward all correct or incorrect responses than would be expected by the binomial distribution, $\chi^2(2, N = 432) = 39.70$, $p < .001$.

According to the structure imposition account, this pattern of responding should also hold in the two recognition blocks. As predicted, the distribution of item responses for the two recognition blocks (135 consecutive corrects vs. 105.51 expected, 157 one correct/one error vs. 215.97 expected, and 140 consecutive errors vs. 110.51 expected) was more skewed toward all correct or incorrect responses than would be expected by the binomial distribution, $\chi^2(2, N = 432) = 32.21$, $p < .001$.

SUSTAIN and RULEX also predict which regular feedback items will most likely be characterized as consistent or inconsistent. Although there is little structure to characterize Categories A and B, Items A1 and B1 in Table 8 serve as the prototypes of their categories. RULEX predicts these items are more likely to follow a constructed rule. Likewise, SUSTAIN predicts that these items are more likely to cluster with other items from the same category. Therefore, both models predict fewer learning errors for these items. As predicted, prototype items resulted in significantly higher classification accuracy (.74 vs. .64) than the other regular feedback items, $t(71) = 4.91$, $p < .001$. The models also predict poorer recognition for the prototype items than for the other regular feedback items. Though the effect (.45 vs. .51) did not reach significance, $t(71) = 1.38$, $p \approx .17$, it was in the predicted direction.

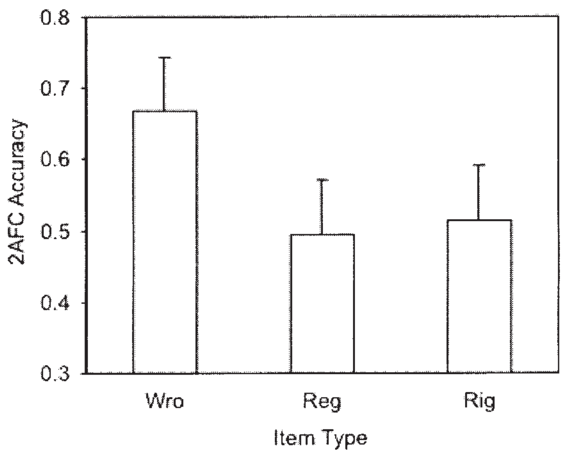

Figure 3. Mean accuracies in the recognition phase are shown along with 95% within-subject confidence intervals (Loftus & Masson, 1994). Error bars represent standard errors of the mean. Wro = the always-wrong item; Reg = the regular feedback items; Rig = the always-right item; 2AFC = two-alternative forced choice.

Model Analyses

The model fits for the recognition data are shown in Table 9. SUSTAIN (three parameters, RMSD = 0.011) correctly predicts the basic results, and its estimates fall within the confidence intervals shown in Figure 3. RULEX (five parameters, RMSD = 0.076) correctly predicts the recognition advantage of the always-wrong item, but underpredicts recognition for the always-right item. The context model (parameter invariant, RMSD = 0.095) incorrectly predicts uniform performance across item types, leading to underpredicting recognition for the always-wrong item. The context model's predictions are parameter invariant because all items bear the same similarity relations to one another. Because RULEX and the context model have difficulty with different aspects of the data, the combined model (six parameters, RMSD = 0.049) fairs better than its constituent models.

Discussion

Whereas Experiments 1 and 2 focused on the role of knowledge structures in directing encoding and recognition, Experiment 3 adopted irregular category structures and instead manipulated the error rate of items during learning. The main finding from Experiment 3 was that the always-wrong item, which always resulted in an error during the learning phase, was better remembered than other items that resulted in fewer errors. However, further analysis of the data revealed that the effect of errors on recognition was likely mediated by structural and attentional mechanisms.

For instance, response time increased as error rate increased in the learning phase. One reasonable interpretation is that subjects allocated more resources to encoding items that were difficult to classify (cf., Stern et al., 1984). Perhaps more interesting are analyses that support predictions made by SUSTAIN and RULEX concerning subjects' imposition of structure on Experiment 3's irregular category structures. These analyses suggest that subjects treated stimulus items as being consistent or inconsistent with an imposed abstraction (e.g., rule, cluster, or schema). Consistent items result in fewer errors during learning and worse recognition than with inconsistent items. These models and analyses suggest that the role of errors in determining recognition memory is mediated through imposed knowledge structures. One possible conclusion is that the drive to extract meaning or abstractions from examples renders it impossible to study the effect of errors on recognition memory independent of structural considerations.

These effects of knowledge structures are somewhat surprising, given how few exemplars were used and how poorly structured the categories were. Others have found that exemplar strategies dominate with such situations (Minda & Smith, 2001; J. D. Smith & Minda, 1998). Indeed, the effects of imposed structure were subtler than in previous experiments. Still, the results from Experiment 3 were best accounted for by SUSTAIN, which posits schemalike structures to represent knowledge. In contrast, the context model was unable to account for the results.

General Discussion

Experiments 1–3 demonstrate that work from the memory literatures is relevant to understanding category learning. These literatures, along with our results, suggest that items are better remembered to the extent that they are differentiated from a salient knowledge structure. The results from Experiments 1–3 are problematic for current models of category learning and recognition. The findings suggest new directions for categorization, schema, stereotype, and memory research. Following a brief review of our results, we discuss possible points of integration among these literatures.

Overview of Empirical Results

Experiment 1 manipulated the salience of knowledge structures to test the prediction that items differentiated from a more salient knowledge structure are remembered better (cf. Rojahn & Pettigrew, 1992). Knowledge structures took the form of imperfect rules, and saliency was manipulated by the number of items satisfying the rule (cf. Koffka, 1935). The results supported our predictions and followed from work in the memory literatures, suggesting that mental representations of regularities are more schemalike than rulelike.

To dissociate violation of structure from making errors, Experiment 2 extended our analysis to the domain of unsupervised

Table 9
Two-Alternative, Forced-Choice Recognition Performances Observed in Experiment 3 (With 95% Confidence Intervals) and Predicted by the Models

Item	Obs	Con	RUL	C + R	SUS
Wro	.67 ± .08	.56[a]	.62	.65	.67
Reg	.49 ± .08	.56	.50	.55	.49
Rig	.51 ± .08	.56	.42[a]	.48	.50

Note. Item types included always-wrong item (Wro), regular-feedback items (Reg), and always-right item (Rig). Obs, Con, RUL, C + R, and SUS stand for observed data, context model, RULEX (rule-plus-exception model), combined model, and SUSTAIN (supervised and unsupervised stratified adaptive incremental network model), respectively.
[a] The predicted value falls outside of the confidence intervals.

learning with a category structure akin to that of natural categories (Rosch & Mervis, 1975). Outside of the laboratory, unsupervised learning is probably more prevalent than classification learning (Love, 2003b). In contrast to Experiment 1, the category structure was not defined by experimenter-supplied feedback but was specified by dimension-value frequency and correlations among stimulus dimensions. The learning phase was reminiscent of the observational procedures used in schema, stereotype, and other memory research. The results from Experiment 2 suggested that subjects formed a schemalike knowledge structure to encode the items that conformed to a common statistical pattern and encoded the deviant item that violated the pattern against this backdrop. Mirroring the results in supervised learning, the deviant item was remembered better than the schema-consistent items.

Work in the schema and stereotype literatures has focused on how knowledge structures guide encoding and retrieval. Typically, errors accompany the violation of a salient knowledge structure during learning. Instead of focusing on category structure, Experiment 3 manipulated the error rate to examine whether errors per se could lead to improved recognition. The results indicated that errors are a correlate, but not a cause of enhanced recognition. Subjects imposed structure on the relatively unstructured categories of Experiment 3. Items that violated these imposed structures were better remembered. The results also suggested that errors may correlate with enhanced recognition because of the extra study time devoted to error-prone items (cf. Metcalfe, 2002; Thiede & Dunlosky, 1999).

Taken together, the results from Experiments 1–3 argue that the drive to discover structure is pervasive. Once a structure is discovered or imposed, it guides the encoding of subsequent experiences. Items are better remembered to the extent that they violate a salient knowledge structure. The more entrenched the knowledge structure, the better memory is for items differentiated from it. Other explanations, although perhaps accounting for aspects of the data, do not fare as well. Errors did not differentiate between the two types of deviant items in Experiment 1, played no role in Experiment 2, and appeared to be mediated through knowledge structures and study time allocation in Experiment 3. Explanations based on study time also failed to explain the entire pattern of results.

Overview of Simulations

The context model, RULEX, their combination (i.e., the combined model), and SUSTAIN were fit to the recognition data. Overall, SUSTAIN fit the data the best, capturing the quantitative and qualitative patterns of results for all studies. The other models had trouble accounting for aspects of the data from each of our three experiments.

Out of all the models, SUSTAIN's operation is most consistent with the memory literatures. SUSTAIN's cluster representations are somewhat analogous to schemas or stereotypes. Additional clusters are recruited in response to surprising events, which allows SUSTAIN to differentiate inconsistent items. This differentiation allows SUSTAIN to predict a recognition advantage for inconsistent information.

RULEX, a hypothesis-testing model that stores exceptions to rules, can also predict a recognition advantage for inconsistent information in a number of circumstances. Both RULEX and SUSTAIN predict that subjects impose structure on learning problems and that the effect of errors on recognition is mediated through these imposed structures. However, RULEX's rule-based representations of knowledge are insensitive to structure saliency (e.g., frequency), are too abstract (e.g., studied rule-following items and novel items are treated similarly), and are not readily applied to unsupervised learning. While capturing aspects of the results, these departures from schema-inspired category representations prevented RULEX from completely accounting for the results of any study.

The context model diverged the farthest from the schema-inspired representations and also faired the worst in terms of fitting the data. The context model is an exemplar model and stores every training item, leaving no role for knowledge structures to guide encoding. The context model could not predict that items that deviate from a regularity are better remembered. The combined model, which generates recognition judgments by pooling RULEX's and the context model's outputs, addressed some of RULEX's shortcomings in regard to having overly abstract representations. The exemplar representations of the context model allowed the combined model to differentiate between rule-following and novel items. However, the combined model, despite its numerous parameters, did not account for all aspects of the data.

Further Synergies and Future Research

Inference learning. Whereas the classification learning tasks considered in Experiments 1 and 3 stress information that discriminates between categories, other induction tasks, such as inference learning, stress the internal structure of the individual categories (Chin-Parker & Ross, 2002; Yamauchi, Love, & Markman, 2002; Yamauchi & Markman, 1998). In inference learning, the learner is provided with the category label and predicts the value of an unknown perceptual dimension. For example, the learner might predict the color of a stimulus given the other perceptual dimensions and the category membership of the item. On the next trial, he or she might predict the size of the stimulus. Inference and classification learning are informationally equivalent because after receiving feedback, complete stimulus information is available in both induction tasks. Despite this equivalence, dramatically different patterns of learning and transfer are observed (see Markman & Ross, 2003, for a review).

Given that SUSTAIN has successfully accounted for inference and classification learning data (Love, Markman, & Yamauchi, 2000; Love et al., 2004), one natural question is how differences in these induction tasks will affect recognition. SUSTAIN makes some interesting predictions. For example, SUSTAIN (along with our differentiation account of recognition) predicts that acquiring the categories in Experiment 1 through inference learning on the rule dimension will reverse the pattern of observed findings. Under inference learning, the exception in the large category should be remembered better than the exception in the small category. In the case of inference, the mapping is from category membership to rule dimension rather than from rule dimension to category membership as in classification learning. Thus, in inference learning, the exception in the larger category violates the more salient regularity, and the predicted pattern of results for inference learning is the opposite of classification learning.

Cognitive neuroscience. Models of category learning have helped provide a theoretical understanding of category learning and recognition data from neuropsychological studies. For example, amnesic patients' ability to acquire simple categories while showing impaired recognition has been used to argue for the presence of multiple memory systems (Knowlton & Squire, 1993), but this interpretation of the data has been challenged by simulations that demonstrate that this dissociation can be modeled within a single exemplar system (Nosofsky & Zaki, 1998). SUSTAIN successfully accounted for our results by utilizing the same mechanisms for categorization and recognition. SUSTAIN's success, coupled with Occam's razor, disfavors accounts that posit distinct systems for categorization and recognition.

Although category learning models have been useful in providing a theoretical understanding of cognitive neuroscience findings, these findings can also provide helpful constraints on the development of category learning models. Recent work has attempted to relate the operation and parts of SUSTAIN to structures in the brain through simulation of data from infant, amnesic, normal, and aging populations (Gureckis & Love, 2004; Love, 2003a). Initial results suggest that the hippocampus is necessary for cluster recruitment, clusters initially reside in the medial temporal lobe, and SUSTAIN's attentional mechanism is instantiated by frontal loops.

Alignable and nonalignable differences. The findings from Experiments 1–3 may relate to literatures that appear quite distant on the surface. Work in analogy suggests that alignable differences are more effective retrieval cues and play a more important role in preference formation than nonalignable differences (Markman & Gentner, 1997; Zhang & Markman, 1998). Alignable differences arise from somewhat salient common dimensions. For example, an alignable difference of a car and a motorcycle is the number of wheels. In this case, the common dimension is having wheels. Nonalignable differences arise when one object has a dimension that the other object lacks. For example, a seat belt is a nonalignable dimension for a motorcycle and a car because motorcycles do not have the dimension of restraining device.

In relation to our work, the common dimensions that give rise to alignable differences can be viewed as common structures or schemas. Collections of these common dimensions may define a superordinate category that includes both object categories (cf. Markman & Wisniewski, 1997). In effect, these common dimensions define knowledge structures (akin to the rule-relevant dimension in Experiment 1) that highlight mismatches across object categories. In keeping with our results, deviance from these knowledge structures is highlighted.

Final Note

One exciting aspect of this line of research is that it affects work within category learning (as evidenced by the challenges it poses to current models of category learning and subsequent recognition) and exposes connections to work outside of category learning. Although the discussion here has largely focused on the benefits of importing ideas from work in schemas and stereotypes to the category learning literature, the potential benefit to these areas is at least as great. Work in category learning typically offers a great deal of experimental control that is not afforded to research that relies on tapping preestablished knowledge structures. Furthermore, it is often impossible to monitor the development of schemas and stereotypes, let alone direct it. These methodological benefits, paired with computational models that make precise predictions, may prove to be valuable tools for research in schemas and stereotypes. The work presented here is an initial step in exploring the conceptual interconnections between a number of areas that appear concerned with similar issues. Still, across methods and subfields, the basic conclusions from work in category learning, analogy, schemas, and stereotypes appear in concert.

References

Ahn, W. K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16,* 81–121.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

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

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17,* 1–12.

Ashby, F., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61,* 1178–1199.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology, 36,* 203–272.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177–220.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer & R. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within category correlations. *Memory & Cognition, 30,* 353–362.

Craik, F. I., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 649–657.

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

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 224–240.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316–355.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

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

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 3–28.

Goodman, G. S. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology, 12,* 473–495.

Graesser, A. C. (1981). *Prose comprehension beyond the word*. New York: Springer-Verlag.

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 59–109). New York: Academic Press.

Green, R. T. (1956). Surprise as a factor in the von Restorff effect. *Journal of Experimental Psychology, 52,* 340–344.

Gureckis, T. M., & Love, B. C. (2004). Common mechanisms in infant and adult category learning. *Infancy, 5,* 173–198.

Hastie, R. (1981). Schematic principles in human memory. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social Cognition: The Ontario Symposium* (Vol. 1, pp. 39–88). Hillsdale, NJ: Erlbaum.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37,* 25–38.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21,* 107–112.

Heit, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 712–731.

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

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1359–1366.

Ito, M. (1989). Long-term depression. *Annual Review of Neuroscience, 12,* 85–102.

Jenkins, W. O., & Postman, L. (1948). Isolation and "spread of effect" in serial learning. *American Journal of Psychology, 61,* 214–221.

Kamin, L. J. (1969). Selective association and conditioning. In N. J. Mackintosh (Ed.), *Fundamental issues in instrumental learning* (pp. 42–64). Halifax, Nova Scotia, Canada: Dalhousie University Press.

Knowlton, B. J., & Squire, L. R. (1993, December 10). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science, 262,* 1747–1749.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Kopp, B., & Wolff, M. (2000). Brain mechanisms of selective learning: Event-related potentials provide evidence for error-driven learning in humans. *Biological Psychology, 51,* 223–246.

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

Levine, D. (1996). *Users guide to the PGAPack parallel genetic algorithm library* (Tech. Rep. No. ANL-95/18). Argonne, IL: Argonne National Laboratory.

List, J. A. (1986). Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology, 22,* 50–57.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565–572.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Love, B. C. (2003a, September). *Infants, amnesiacs, and the MTL*. Paper presented at the J. S. McDonnell Foundation Cognitive Neuroscience of Category Learning Meeting, New York.

Love, B. C. (2003b). The multifaceted nature of unsupervised category learning. *Psychonomic Bulletin & Review, 10,* 190–197.

Love, B. C., Markman, A. B., & Yamauchi, T. (2000). Modeling classification and inference learning. *Proceedings of the Fifteenth National Conference on Artificial Intelligence, 15,* 136–141.

Love, B. C., & Medin, D. L. (1998). Modeling item and category learning. In S. J. Derry & M. A. Gernsbacher (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 639–644). Hillsdale, NJ: Erlbaum.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of human category learning. *Psychological Review, 111,* 309–332.

Love, B. C., & Sakamoto, Y. (2003). Schematic influences on category learning and recognition memory [Abstract]. *The 44th Annual Meeting of the Psychonomic Society, 8,* 26.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimulus with reinforcement. *Psychological Review, 82,* 276–298.

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

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. *Psychological Science, 8,* 363–367.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129,* 592–613.

Markman, A. B., & Wisniewski, E. (1997). Similar and different: The differentiation of basic level category. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 54–70.

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.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19,* 242–279.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131,* 349–363.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 775–799.

Murphy, G. L., & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 904–919.

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research, 171,* 213–224.

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

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

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

Nosofsky, R. M., & Zaki, S. F. (1998). Dissociations between categorization and recognition in amnesic and normal individuals. *Psychological Science, 9,* 247–255.

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

Pazzani, M. J. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 416–432.

Pearce, J., & Hall, G. (1980). A model for Pavlovian conditioning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 587–595.

Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology, 69,* 309–315.

Pinker, S. (1991, August 2). Rules of language. *Science, 253,* 530–535.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roediger, H. L., III., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

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

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology, 15,* 343–355.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986, October 9). Learning representations by back-propagating errors. *Nature, 323,* 533–536.

Sara, S. J., & Segal, M. (1991). Plasticity of sensory responses of locus coeruleus neurons in the behaving rat: Implications for cognition. *Progressive Brain Research, 88,* 571–585.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Neuroscience, 1,* 199–207.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 305–354). San Diego, CA: Academic Press.

Sentis, K. P., & Burnstein, E. (1979). Remembering schema-consistent information: Effects of a balance schema on recognition memory. *Journal of Personality and Social Psychology, 37,* 2200–2211.

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

Sherman, J. W., & Frost, L. A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin, 26,* 26–34.

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

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1411–1430.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36,* 941–950.

Spalding, T. L., & Murphy, G. L. (1996). Effects of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 525–538.

Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 440–463.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and developmental literatures. *Psychological Bulletin, 111,* 42–61.

Stern, L. D., Marrs, S., Millar, M. G., & Cole, E. (1984). Processing time and recall of inconsistent and consistent behaviors of individuals and groups. *Journal of Personality and Social Psychology, 47,* 253–262.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social Cognition: The Ontario Symposium* (Vol. 1, pp. 89–134). Hillsdale, NJ: Erlbaum.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024–1037.

von Restorff, H. (1933). Analyse von vorgängen in spurenfeld: I. Über die wirkung von bereichsbildungen im spurenfeld [Analysis of processes in the memory trace: I. On the effect of group formations on the memory trace]. *Psychologische Forschung, 18,* 299–342.

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin, 63,* 410–424.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18,* 158–194.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 585–593.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124–149.

Zhang, S., & Markman, A. B. (1998). Overcoming the early entrant advantage via differentiation: The role of alignable and nonalignable differences. *Journal of Marketing Research, 35,* 413–426.

Appendix

Modeling Details

In this appendix, we first describe the procedures for the model fits. We then describe the context model, RULEX, the combined model, and SUSTAIN.

Model Fit Procedures

All of the models were simulated in a manner as consistent as possible with the procedures used in the original human experiments. For example, only the model runs that reached the criterion applied to humans (e.g., above chance performance in the learning phase) were included in the model analyses. Only the results from the recognition phase were fit because our main interest was in examining recognition memory, and all the models gave reasonable accounts of the learning and transfer phase data.

Fit was measured by root mean squared deviations (RMSD). The best fitting parameters were found by searching the parameter space using the PGAPack genetic algorithm (Levine, 1996). The genetic algorithm searched the parameter space to find the parameter set that minimized RMSD.

Models generated recognition ratings during simulations of Experiment 2 and Palmeri and Nosofsky's (1995) Experiment 3. The models' recognition ratings were based on the familiarity, F, as given for the context model in Equation 4, for RULEX in Equation 5, for the combined model in Equation 8, and for SUSTAIN in Equation 11. In fitting recognition ratings, a linear relationship was assumed between F and the human ratings. The best fitting parameters (slope and intercept) for the models for these simulations are shown in Table A1.

Models made two-alternative forced-choice judgments between a studied and a novel stimulus in simulations of Experiments 1 and 3. The probability of choosing the studied item was determined by the exponential decision function,

$$P(\text{old}) = \frac{e^{rF_{\text{old}}}}{e^{rF_{\text{old}}} + e^{rF_{\text{new}}}}, \tag{A1}$$

where F_{old} is the model's familiarity for the studied stimulus, F_{new} is the model's familiarity for the novel stimulus, and r is the recognition decision parameter. The best fitting values for r are shown in Table A2. Like the linear response function parameters of slope and intercept, the r recognition decision parameter affects the quantitative but not qualitative predictions of the models.

Table A1
The Slope and Intercept Used in the Linear Transform of F (Generated by the Models) Into Recognition Ratings for Experiment 2 (Exp. 2) and Palmeri and Nosofsky's (1995) Experiment 3 (PN 3)

	PN 3		Exp. 2	
Model	Slope	Intercept	Slope	Intercept
Context	183.342	−1,418.536	0.189	0.648
RULEX	1.643	5.459	—	
Combined	2.317	5.159	—	
SUSTAIN	18.137	−7.535	4.645	−1.649

Note. The dashes indicate that RULEX (rule-plus-exception model) and the combined model were not fit to Experiment 2. SUSTAIN = supervised and unsupervised stratified adaptive incremental network model.

Table A2
The Best Fitting Values of the Recognition Decision Parameter r

Model	Exp. 1	Exp. 3
Context	3.329	2.431
RULEX	39.907	2.855
Combined	2.620	49.567
SUSTAIN	44.445	5.322

Note. The r parameter transforms the F values generated by models into forced-choice probabilities in Experiment 1 and Experiment 3 simulations. Exp. = experiment. RULEX = rule-plus-exception model; SUSTAIN = supervised and unsupervised stratified adaptive incremental network model.

Context Model

The context model fully stores each training example. The probability that a given stimulus S_i is classified into Category A is determined by summing the similarity of S_i to all members of Category A and dividing by the summed similarity of S_i to members of both Categories A and B,

$$P(\text{A}|S_i) = \frac{\sum_{j\in \text{A}} s_{ij}}{\sum_{j\in \text{A}} s_{ij} + \sum_{j\in \text{B}} s_{ij}}, \tag{A2}$$

where s_{ij} is the similarity between S_i and S_j determined by the multiplicative rule,

$$s_{ij} = \prod_{m=1}^{M} s_m^{\delta_m(i,j)}, \tag{A3}$$

where s_m is a parameter indicating the similarity of mismatches along dimension m, and δ_m (i, j) is an indicator function equal to 0 if stimulus S_i and S_j match along dimension m and set equal to 1 if they mismatch along dimension m. The similarity parameter, s_m, represents a combination of dimension salience and selective attention.

Recognition decisions in the context model are based on the absolute summed similarity of a stimulus item to the stored exemplars of both categories. The familiarity, F_i, of a stimulus item, S_i, is

$$F_i = \sum_{j\in \text{A}} s_{ij} + \sum_{j\in \text{B}} s_{ij}. \tag{A4}$$

In Experiment 3 and in Palmeri and Nosofsky's (1995) Experiment 3, the summed similarity of a stimulus item to the stored exemplars was the same for all items. Thus, varying s_m had no effect on the context model's predictions. In Experiment 1, subjects were instructed to attend to the rule dimension. One parameter (s_{rule} set to .025) was used for the rule dimension, and another parameter (s_{nonrule} set to .757) was used for the other dimensions. In Experiment 2, only one parameter (s_x set to .011) was needed because all dimensions are structurally equivalent.

RULEX

In the RULEX model, classification learning is based on the acquisition of rules. Rules can be supplemented by the partial storage of exceptions. First, RULEX searches for a perfect single-dimension rule. A dimension is sampled according to W_i (each dimension's intrinsic salience parameter) and a single-dimension rule is formed. When physical dimensions are

(*Appendix continues*)

randomly assigned to abstract dimensions, RULEX assumes equal dimension saliences. When a rule fails, it is discarded and a new dimension is sampled.

If all dimensions fail to provide a perfect single-dimension rule, RULEX searches for an imperfect single-dimension rule through the same process used to search for a perfect single-dimension rule. An imperfect rule is retained for a minimum number of trials set equal to the number of training items by default. When the minimum number of trials is reached, the imperfect rule is retained only if performance exceeds *lax* (the lax criterion parameter). The imperfect rule is evaluated after a given number of trials is set equal to twice the number of training items. At this point, if performance exceeds *scrit* (the strict criterion parameter), the imperfect rule is permanently stored. Otherwise, it is discarded and another dimension is selected. If all dimensions have been sampled, RULEX searches for a conjunctive rule by using a similar process.

Following the permanent storage of a single-dimension or a conjunctive rule, RULEX attempts to store exception items that contradict the stored rule. The dimensions of the exception that contradict the stored rule are sampled with probability one. Other dimensions are probabilistically sampled with probability *pstor* (the probability storage parameter). If a dimension is not sampled, it is stored as a "wildcard" that can match any value. The probability that an exception will be stored in memory is equal to $pstor^N$, where N is the number of dimensions that were sampled in the previous step.

When making a classification decision, RULEX first searches for all stored exceptions that match a given stimulus item. When an item matches a stored exception, the classification response is based on the matched exception. If an item matches multiple exceptions, the response depends on how many of the matched exceptions signal Category A or Category B. An exception that results in a classification error is removed from memory. RULEX applies the rule if no exceptions match the item.

The familiarity of a stimulus, S_i, is determined by its summed similarity to each of the exceptions, X_j, and is given by

$$F_i = \sum_{j \in \text{Exc}} s_{ij}, \tag{A5}$$

where s_{ij} is

$$s_{ij} = \prod_{m=1}^{M} \Theta_m(i,j), \tag{A6}$$

where

$$\Theta_m(i,j) = \begin{Bmatrix} s_w & \text{if } X_j \text{ contains a wildcard on dimension } m \\ s_s & \text{if } S_i \text{ mismatches } X_j \text{ on dimension } m \\ 1 & \text{if } S_i \text{ matches } X_j \text{ on dimension } m \end{Bmatrix}, \tag{A7}$$

where s_w and s_s are the wildcard and mismatch similarity parameters. The best fitting parameter values for RULEX are shown in Table A3.

Table A3
The Best Fitting Parameters for RULEX

Parameter	PN 3	Exp. 1	Exp. 3
Rule dimension weight (W_{rule})		1.000	
Lax criterion (*lax*)	0.602	0.624	0.300
Strict criterion (*scrit*)	0.750	0.413	0.112
Probability storage (*pstor*)	1.000	0.911	1.000
Wildcard similarity (s_w)	0.004	0.199	0.153
Mismatch similarity (s_s)	0.130	0.975	0.708

Note. PN 3 = Palmeri and Nosofsky's (1995) Experiment 3; Exp. = experiment; RULEX = rule-plus-exception model.

Combined Model

Classification decisions in the combined model are based on RULEX, but recognition decisions are based on a combination of RULEX's response (discussed previously) and residual exemplar memory (implemented using the context model). The familiarity of a stimulus, S_i, is defined by

$$F_i = \omega F_i^{X} + (1 - \omega) F_i^{R}, \tag{A7}$$

where F_i^X is the summed similarity of S_i to the exceptions (familiarity as defined in Equation 5 for RULEX), F_i^R is the residual summed similarity of S_i to all exemplars (familiarity as defined in Equation 4 for the context model), and the parameter ω weights the contributions of these two sources of familiarity. Setting ω equal to 0 results in a version of the context model (with equal dimension similarity, s, for all dimensions), whereas setting ω equal to 1 is equivalent to using RULEX alone. The best fitting parameter values for the combined model are shown in Table A4.

SUSTAIN

At the start of training, SUSTAIN creates a single cluster centered on the first training item. SUSTAIN assigns subsequent training items to the most similar cluster. Cluster recruitment is triggered by a surprising event. In supervised learning, a surprising event is a prediction error. In unsupervised learning, a surprising event is encountering a sufficiently novel stimulus. When items are successfully assigned to a cluster, the winning cluster adjusts its position to move toward its newest member, strengthens its association weights, and adjusts attention to favor predictive dimensions. Items are classified based on cluster assignment. Recognition strength is determined by summing the output of all clusters.

A nominal stimulus dimension containing k distinct values is represented by k input units. The unit that denotes the value of the dimension is set to one, and all of the other units forming the dimension are set to 0. A complete stimulus is represented by $I^{\text{pos}_{ik}}$, where i indexes the stimulus dimension, and k indexes the nominal values for dimension i. The pos in I^{pos} denotes that the current stimulus occupies a particular position in a multidimensional representational space.

The distance μ_{ij} between the ith stimulus dimension and cluster j's position along the ith dimension is

$$\mu_{ij} = \frac{1}{2} \sum_{k=1}^{v_i} |I^{\text{pos}_{ik}} - H_j^{\text{pos}_{ik}}|, \tag{A9}$$

where v_i is the number of different nominal values on the ith dimension, $I^{\text{pos}_{ik}}$ is the position of the input stimulus on the ith dimension for value k, and $H^{\text{pos}_{ik}}$ is cluster j's position on the ith dimension for value k. The distance μ_{ij} is always between 0 and 1 inclusive.

The activation of a cluster is given by

$$H_j^{\text{act}} = \frac{\sum_{i=1}^{m} (\lambda_i)^{\gamma} e^{-\lambda_i \mu_{ij}}}{\sum_{i=1}^{m} (\lambda_i)^{\gamma}}, \tag{A10}$$

where H_j^{act} is the activation of the jth cluster, m is the number of stimulus dimensions, λ_i is the tuning of the receptive field (which implements attention in SUSTAIN) for the ith input dimension, and γ is the attentional parameter (always nonnegative).

Clusters compete to respond to input patterns and in turn inhibit one another,

$$H_j^{\text{out}} = \frac{(H_j^{\text{act}})^{\beta}}{\sum_{i=1}^{n} (H_i^{\text{act}})^{\beta}} H_j^{\text{act}}, \tag{A11}$$

Table A4
The Best Fitting Parameters for the Combined Model

Parameter	PN 3	Exp. 1	Exp. 3
Rule dimension weight (W_{rule})		0.917	
Lax criterion (*lax*)	0.631	0.592	0.602
Strict criterion (*scrit*)	0.628	0.444	0.166
Probability storage (*pstor*)	0.892	0.987	0.957
Wildcard similarity (s_w)	0.020	0.433	0.551
Mismatch similarity (s_s)	0.181	0.756	0.425
Exemplar similarity (s)	—	0.235	—
Exception weight (ω)	0.707	0.799	0.822

Note. A dash indicates that all values yield the same fit. PN 3 = Palmeri and Nosofsky's (1995) Experiment 3; Exp. = experiment.

where n is the number of the clusters, and β is the lateral inhibition parameter (always nonnegative) that regulates cluster competition. SUSTAIN's recognition responses are determined by summing H_j^{out} for all clusters. When classifying an item, SUSTAIN selects the winning cluster by setting H_j^{out} for all nonwinning H_j 0.

Activation is spread from the clusters to the output units of the unknown (queried) dimension z by

$$C_{zk}^{\text{out}} = \sum_{j=1}^{n} w_{j,zk} H_j^{\text{out}}, \tag{A12}$$

where C_{zk}^{out} is the output of the output unit representing the kth nominal value of the unknown zth dimension, n is the number of clusters, and $w_{j,zk}$ is the weight from cluster j to output unit C_{zk}. In classification learning, z is the category label.

The probability of making a response k *(the k*th nominal value) for the queried dimension z is

$$P(k) = \frac{e^{(d \cdot C_{zk}^{\text{out}})}}{\sum_{j=1}^{v_z} e^{(d \cdot C_{zk}^{\text{out}})}}, \tag{A13}$$

where d is the response parameter (always nonnegative), and v_z is the number of nominal units (and hence output units) forming the queried dimension z.

After responding, feedback is provided to SUSTAIN. In classification learning, a new cluster is recruited when the winning cluster predicts the incorrect category label. In unsupervised learning, a new cluster is recruited when the activation of the most activated cluster is below the parameter τ. Recruited clusters are centered on the misclassified example (i.e., all μ_{ij} will be 0 for the new cluster and the current stimulus).

For the winning cluster H_j, the position of the cluster is adjusted by

$$\Delta H_j^{\text{pos}_{ik}} = \eta(I^{\text{pos}_{ik}} - H_j^{\text{pos}_{ik}}), \tag{A14}$$

where η is the learning rate parameter. The winning cluster moves toward the current stimulus. This learning rule centers the cluster amid its members.

Receptive field tunings (which implement dimensional attention) are updated according to

$$\Delta\lambda_i = \eta e^{-\lambda_i \mu_{ij}}(1 - \lambda_i \mu_{ij}), \tag{A15}$$

where j is the index of the winning cluster. Only the winning cluster updates the value of λ_i.

The one layer delta learning rule (Rumelhart et al., 1986) is used to adjust weights from clusters to output units and is given by

$$\Delta w_{j,zk} = \eta(t_{zk} - C_{zk}^{\text{out}})H_j^{\text{out}}, \tag{A16}$$

where z is the queried dimension. Note that only the winning cluster will have its weights adjusted because it is the only cluster with a nonzero output.

In Experiment 1, subjects were alerted to the rule dimension. The $\lambda_{\text{distinct}}$ parameter was used to allow SUSTAIN to initially weight the rule dimension more than the nonrule dimensions (λ is normally set to 1). The best fitting parameter values for SUSTAIN are shown in Table A5.

Table A5
The Best Fitting Parameters for SUSTAIN

Parameter	PN 3	Exp. 1	Exp. 2	Exp. 3
Attentional focus (γ)	0.614	2.799	4.159	1.798
Cluster competition (β)	10.144	0.510	9.930	9.044
Decision consistency (d)	33.616			
Learning rate (η)	0.063	0.076	0.044	0.070
Distinct focus ($\lambda_{distinct}$)		5.346		
Recruitment threshold (τ)			0.291	

Note. The parameter d for Experiment 1 and Experiment 3 was yoked to the recognition decision parameter r. PN 3 = Palmeri and Nosofsky's (1995) Experiment 3; Exp. = experiment; SUSTAIN = supervised and unsupervised stratified adaptive incremental network model.

Received April 1, 2003
Revision received May 10, 2004
Accepted May 21, 2004 ■