The Origin of Exemplar Effects in Rule-Driven Categorization

Guy L. Lacroix
Concordia University

Serge Larochelle
Université de Montréal

Gyslain Giguère
Université du Québec à Montréal

S. W. Allen and L. R. Brooks (1991) have shown that exemplar memory can affect categorization even when participants are provided with a classification rule. G. Regehr and L. R. Brooks (1993) argued that stimuli must be individuated for such effects to occur. In this study, the authors further analyze the conditions that yield exemplar effects in this rule application paradigm. The results of Experiments 1–3 show that interchangeable attributes, which are not part of the rule, influence categorization only when attention is explicitly drawn on them. Experiment 4 shows that exemplar effects can occur in an incidental learning condition, whether stimulus individuation is preserved or not. The authors conclude that the influence of exemplar learning in rule-driven categorization stems from the attributes specified in the rule or in the instructions, not from the stimulus gestalts.

Exemplar effects are robust in the inductive category-learning paradigm (Medin, Altom, Edelson, & Freko, 1982; Medin & Schaffer, 1978; Nosofsky, 1986, 1988). However, the category-construction literature shows that participants use simple and typically one-dimensional rules to sort items and that the influence of interstimulus similarity on sorting behavior is quite limited (Ahn & Medin, 1992; Regehr & Brooks, 1995). Thus, before Allen and Brooks’s (1991) work, it was not clear to what extent exemplar memory would develop and influence categorization when participants can also rely on an explicit rule. The finding that exemplar memory does influence performance in rule-driven categorization was noteworthy and has since been reported as an empirical example of the influence of exemplar memory on categorization in many reviews (for instance, Goldstone, 1994; Goldstone & Bar-salou, 1998; Hahn & Chater, 1998; Murphy, 2002).

Allen and Brooks’s (1991) exemplar effects were shown using an ingenious paradigm, which we adopted in our experiments. The stimuli were individuated fictional animals made of five binary attributes. Table 1 shows the attribute values of the stimuli used in the experiments reported here. The participants categorized the stimuli using a classification rule that was stated at the onset of the experiment. It required the animals to have at least two of three prespecified attribute values to belong to a given category. Otherwise, the animals belonged to the opposite category. In our experiments, the three leftmost attributes in Table 1 (tail type, back pattern, and head shape) were stated in the rule. By assuming that all values of 1 in the table correspond to the attribute values specified in the rule, the reader can verify that Training Items 1–4 belong to one category, whereas Items 5–8 belong to the other. These three attributes were not independently sufficient to determine category membership. The cue validity of each individual attribute stated in the rule was 75%. The two other attributes that were not part of the rule (body type and color in Table 1) were nondiagnostic; each of their values occurred equally often in both categories.
Two types of transfer items were presented in a subsequent test phase. We produced the transfer items by changing the value of one of the rule attributes (the back pattern in Table 1). This manipulation made it possible to have certain transfer items belong to a different category from that of the most similar training item. Allen and Brooks (1991) called such stimuli negative match items. By contrast, transfer items that belonged to the same category as their most similar training item were called positive match items.

Table 1 shows match items next to their corresponding training items, called positive and negative old items. Allen and Brooks found that categorization of negative match items took longer and was more error-prone. For instance, in their Experiment 1, the error rates were 45% for negative match items compared with approximately 20% for both old and positive match items. Response times were close to 1,600 ms for negative match items compared with 1,200–1,400 ms for the other types of items. The authors' interpretation of this negative match effect was that the exemplars seen during the learning phase had been memorized and that, subsequently, the similarity between the training and the transfer items caused the memorized exemplars to be retrieved and to influence the categorization process, despite the availability of a deterministic rule.

In later work, Regehr and Brooks (1993) attempted to specify the conditions that are necessary for exemplar effects to obtain in the rule application paradigm. Hence, they analyzed the effects of individuality at the local (attribute) and global (whole stimulus) levels, varying both factors over experiments. Attribute individuality was obtained when each attribute was implemented in a unique, idiosyncratic way. For example, if a stimulus had the attribute long neck, all stimuli with this attribute value had a physically different long neck (see Figure 5 for examples). The attributes had no individuality when there was a one-to-one correspondence between an attribute value and its physical implementation (see Figure 1 for examples). Individuality at the global level

---

**Table 1**

*Abstract Description of the Stimuli Used in Experiments 1–4*

<table>
<thead>
<tr>
<th>Exemplars</th>
<th>Tail type</th>
<th>Back pattern</th>
<th>Head shape</th>
<th>Body type</th>
<th>Color</th>
<th>Exemplars</th>
<th>Tail type</th>
<th>Back pattern</th>
<th>Head shape</th>
<th>Body type</th>
<th>Color</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Negative</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Positive</td>
<td>17</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Negative</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Note. 0 and 1 represent the binary values for each attribute. Category membership was determined by the rule stated at the onset of the experiment. Assuming that 1 represents the value specified in the rule for tail type, back pattern, and head shape, then Training Items 1–4 (and Transfer Items 9–12) belonged to one category, whereas Training Items 5–8 (and Transfer Items 13–16) belonged to the opposite category. Backgrounds were used only in Experiment 3, in which number of legs replaced colors as a nonrule attribute, and four different colors served as background. For Experiment 4, the attributes were (from left to right): number of legs, spots, body type, neck length, and tail length.

---

Figure 1. Grayscale examples of the stimuli used in Experiments 1 and 2. The condition labels apply to participants for whom the stated rule was a cane-shaped tail, a striped back pattern, and a parabolic head shape (see Table 1). The stimuli used in Experiment 3 were similar except that the nonrule attribute color was changed to number of legs.
was obtained when the composition of the attributes made each stimulus look distinct. Global individuality was absent when interstimulus similarity was high. The results showed that only stimuli that had individuality at both the attribute and global levels produced a negative match effect on error rates and response times (Experiments 1D and 3A). When the stimuli were individuated at the global level but not at the attribute level, a negative match effect was found for response times only (Experiment 2A). Finally, when the stimuli were made of nonindividuated or interchangeable features, negative match effects were not found for error rates or response times (Experiments 1D and 2B). This led Regehr and Brooks to emphasize that it was necessary for the stimuli to form distinctive wholes, gestalts, in order for exemplar effects to occur.

The negative match effects that Allen and Brooks (1991) and Regehr and Brooks (1993) obtained suggest that exemplar memory may exert great influence on classification behavior. In the rule application paradigm, categorization can be performed without any memory of the stimuli, because a perfectly predictive rule is known. Hence, to obtain a negative match effect, exemplar learning must occur incidentally. Moreover, learning must be extensive. Because the categorization rule included only three of the five attributes, one can perform the categorization task without paying attention to the two remaining attributes. However, these two nonrule attributes, which have no cue validity, must also contribute to the memory trace, because they are critical in determining the similarity between the training and corresponding transfer stimuli responsible for the exemplar effects reported. Allen makes this claim very explicitly in an unpublished study of the rule application paradigm with children (Wagner & Allen, 1995): “the perceptual system normally encodes both the diagnostic and nonrule features of a stimulus array and the record that is formed contains both types of information” (p. 4). Ultimately, the idea goes back to Brooks’s (1978) initial conception of nonanalytical processes: “the category membership of an item is inferred from its overall similarity to a known individual or low-level cluster of individuals, where similarity is judged on the basis of aspects or configurations of the stimulus that are not weighted for their criteriality for the particular concept being considered” (p. 180).

If this explanation of the negative match effect is correct, one might have expected exemplar memory to influence the application of the categorization rule whether the stimulus attributes were individualized or interchangeable. Paramount to Brooks and his colleagues’ position (Allen & Brooks, 1991; Regehr & Brooks, 1993) is the assumption that it is the memory trace of the combination of all the attributes of a given stimulus that interferes with the application of the rule. Thus, there is no obvious reason why this task should create a lesser burden for memory when the attributes are individuated. However, if one were to assume that the memory trace of a single attribute, or a small subset of attributes, were sufficient to create such transfer effects, then the importance of individualized attributes would be clear.

In this study, we reexamine under which conditions exemplar effects obtain in the rule application paradigm. In Experiments 1–3, the stimuli were composed of attributes that were not individualized. The stimuli used in Experiment 4 were composed of idiosyncratic features. The reason for focusing on these two extreme conditions is that they impose a very different burden on exemplar learning and memory. In the absence of idiosyncratic attributes, the only information that is unique to each stimulus and that can be used to distinguish the stimuli from one another is the combination of the attribute values. Therefore, many attribute values have to be learned and ultimately recalled to individuate exemplars and to produce exemplar effects. By contrast, when all attribute values are idiosyncratic, one can distinguish exemplars by learning and recalling the value of only one attribute. We will show that the learning that occurs in the context of categorization by rule does not encompass all stimulus attributes or the gestalt resulting from their combination even when the stimuli are highly individuated. Rather, our results show that the exemplar effects obtained in the rule paradigm are caused by a few attributes to which attention is explicitly drawn.

Experiment 1

In Experiment 1, we attempted to obtain negative match effects using stimuli made of interchangeable attribute values. We hypothesized that Regehr and Brooks’s (1993) failure to obtain such effects (in Experiments 1D and 2B) could have been because of an insufficient amount of training, because previous research has shown that longer periods of practice favor exemplar-based classification (Smith & Minda, 1998). In Regehr and Brooks’s experiments, participants saw the eight training items only five times each. In order to determine the effects of practice, our participants were tested after 40 and 160 trials of training. In the longer training condition, the participants viewed the eight training items 20 times each, which gave them an increased learning opportunity.

Nevertheless, we cannot assume that improved memory for the exemplars will necessarily result in an increased negative match effect. One cannot ignore the possibility that rules may also undergo changes with practice. For instance, Anderson (1976, 1983, 1993) has argued that, at early stages of skill acquisition, rules are stored in a declarative format that requires search and interpretation, resulting in slow execution. With practice, some rule elements may be chunked and tested together, and the rules themselves may be stored in a much more readily executable format, leading to some form of automatization. In order for negative match effects to arise in the rule application paradigm, the time taken to access exemplar memory (and the associated category names) would have to be shorter than the time taken to apply the rule. Thus, to assess exemplar memory independently, we submitted participants to an explicit recognition test after completing the learning and transfer phases.

Another possible explanation for Regehr and Brooks’s (1993) failure to obtain negative match effects with stimuli made of interchangeable features is the lack of saliency of the nonrule attributes. As these attributes were peripheral (neck length and number of legs), they may not have been salient enough to attract attention. By contrast, the nonrule attributes used here were either central (body type) or spanned the entire stimulus (color). We did this to ensure that the nonrule attributes would fall within the participants’ visual span while applying the rule. Moreover, the nonrule attributes were selected to be more salient than the rule attributes. In a category-construction task involving similar stimuli, Lacroix and Larochelle (2000) found that 30% of the participants sorted the items on the basis of color, 45% on the basis of

---

1 We thank Jean-Pierre Thibault for useful discussions on this issue.
body shape, and 25% on the basis of one of the remaining attributes (texture, foot type, neck type, and dorsal patterns). Thus, although the rule and the nonrule attributes involved interchangeable values, stimulus composition was deliberately biased in favor of the nonrule attributes at the perceptual level.

**Method**

**Participants**

Thirty-two students at the Université de Montréal, Montréal, Québec, Canada, participated in the study. All received Cdn$3 as compensation for their time. This experiment and all the others reported here were conducted in French.

**Material**

As illustrated in Figure 1, the stimuli were drawings of fictional animals built from five binary attributes: tail type (cane-shaped or stair-shaped), back pattern (striped or spotted), head shape (parabolic or oval), body type (oval or parallelogram), and color (yellow or green). The values that each of the attributes could take were perceptually identical in all stimuli. The three rule attributes determined category membership. They were tail type, back pattern, and head shape. The rule required that at least two of three attribute values specified be present in a stimulus for it to belong to a given category. For example, one instantiation of this rule stated that if a creature had two or three of the following attributes: a cane-shaped tail, stripes, and a parabolic-shaped head, then it belonged to the Tremblay category. Otherwise, it belonged to the Beaulieu category. The two attributes that were not part of the rule, body type and color, were nondiagnostic. They maximized the similarity between the old and corresponding match items (see Table 1 for the abstract structure of the stimuli).

With three binary attributes, there are eight possible rules. All were used with different subgroups of participants. These different rules change the category membership of the stimuli and whether they are positive or negative items, but they do not determine items serving as training versus transfer items. As a consequence, the Exemplar Sets 1–8 and 9–16 served in turn as old items and match items. Our purpose in using all the rules was to insure that all items served in all categories and in all conditions of the experiment. The rules were given to participants in one of two orders: tail, back pattern, and head or head, back pattern, and tail. These two orders avoid a possible bias in the application of the rule. If one were to test the attributes following the order specified in the rule, then one would reach a decision after an equal number of feature tests (i.e., two or three) for both positive and negative items.

For the training and test phases, we chose the 16 items with the same abstract structure as the one in Allen and Brooks (1991). Each transfer item was matched with a training item on every attribute except back pattern. The 16 remaining items that can be constructed from combinations of five binary attributes were used in a recognition test. In all, five stimulus types were produced: positive old (training item whose corresponding transfer item belongs to the same category); negative old (training item whose corresponding transfer item belongs to the opposite category); positive match (transfer item whose corresponding training item belongs to the same category); negative match (transfer item whose corresponding training item belongs to the opposite category); and new item (item seen only in the recognition test that is neither an old nor a match item). All items appeared on a black background. Examples of the stimuli are given in Figure 1.

**Procedure**

The participants were tested individually. All instructions and stimuli were presented on 14-in. video graphics array monitors connected to PC-compatible computers with 80386, 80486, or Pentium Intel microprocessors. The program MEL Professional v.2.01 (Schneider, 1989) was used to give the experimental instructions, present the material, and record the participants’ answers and response times. This aspect of the procedure was identical for all experiments.

In the first phase of the experiment, the participants were given the categorization rule and were instructed to classify the stimuli accordingly. Participants were given 40 trials divided into five blocks. Each block involved the presentation of the eight old items in random order. All trials proceeded as follows. First, a fixation point appeared in the center of the screen for 1,500 ms, immediately followed by an old item that was also centered on the screen. Participants had to categorize the stimulus as quickly as possible while being as accurate as possible (as was requested in Regehr & Brooks’s (1993) experiments). Participants responded by pressing the appropriate key on the keyboard: A for Beaulieu or L for Tremblay. The stimulus remained on display for 2,000 ms after the response in order for feedback to be given. For correct answers, this feedback was the category name. For wrong answers, a short buzzing tone accompanied the correct category name. Once the 2,000 ms elapsed, the screen was cleared. The interstimulus interval was 1,000 ms.

The second phase of the experiment was a test phase. The four positive old, the four negative old, the four positive match, and the four negative match items were presented once each in random order, as done in Regehr and Brooks’s (1993) experiments. This phase proceeded like the previous training phase, except that the stimuli disappeared as soon as a response was made and that no feedback was given. The first two phases of this experiment replicate the basic design of Regehr and Brooks’s experiments.

In the third phase of the experiment, the participants were given an additional 15 blocks (120 trials) of training. They continued to classify the creatures according to the rule. This additional training gave participants more exposure to the exemplars and more practice at applying the rule. The fourth phase of the experiment was identical to the second phase. It provided a second opportunity to test for negative match effects.

The last phase of the experiment was an explicit recognition memory test. Participants were shown the 8 old items, the 8 match items, and the 16 new items in a random order. Their task was to determine whether each particular item had been seen in the training phases of the experiment (identified as those parts of the experiment in which you were given feedback). Hence, this direct memory test forced participants to explicitly discriminate old items, on the one hand, from match and new items, on the other. Responses were given by pressing the 1 key for old items and the space bar for the other types of items. No feedback on accuracy was given.

**Results**

**Classification Task**

In all experiments reported here, we performed separate analyses of variance (ANOVA)s on the error rates and on the mean response times. The data were always filtered prior to the response time analyses. We excluded error trials and trials yielding response times further than three standard deviations away from each individual participant’s mean from the response time analyses. When the elimination of such trials created empty cells in the design, we dropped the implicated participants from the analyses. Following Allen and Brooks (1991), we also always analyzed response times after subjecting the data to a log transformation. In general, we report the results of the ANOVA performed on the raw response times. However, we mention the analyses on log transforms when the results of the two analyses diverge. This occurred only in Experiment 3.

In Experiment 1, the analyses involved $(2) \times (2) \times (2)$ designs with three within-subjects factors: item type (old vs. match), item
value (positive vs. negative), and number of training blocks (5 blocks vs. 20 blocks). In such a design, a negative match effect manifests itself by an Item Type × Value interaction; the difference between negative match and negative old items must be greater than the difference between positive match and positive old items. No participant was lost or excluded from the analyses because of the filtering of errors and excessively long trials.

**Error rates.** The error data pertaining to item type, value, and number of training blocks are presented in the top portion of Figure 2. As can be seen, error rates were very low (all below 4%). Neither the Item Type × Item Value × Number of Training Blocks interaction, $F(1, 31) = .39, MSE = 0.003, p = .54$, nor the Item Type × Value interaction, $F(1, 31) = 0, MSE = 0.004, p = 1$, was significant. Error rates were similar for old and match items, whether positive or negative. However, participants made significantly more mistakes categorizing negative items (2%) than they did categorizing positive items (.02%), $F(1, 31) = 8.9, MSE = 0.003, p = .005$.

**Response times.** The mean response times are shown in the bottom portion of Figure 2. The Item Type × Item Value × Number of Blocks interaction again failed to be significant, $F(1, 31) = .84, MSE = 22.754, p = .37$, and so did the Item Type × Value interaction, $F(1, 31) = .38, MSE = 12.715, p = .54$. After five blocks, response times for negative match items were 1,196 ms versus 1,204 ms for negative old items. Participants were significantly slower in categorizing negative items (1,056 ms) compared with positive items (953 ms), $F(1, 31) = 16, MSE = 42.509, p < .001$.

Hence, we did not succeed in producing a negative match effect. Moreover, some of the differences we did find between positive
and negative items are difficult to explain from an exemplar-based perspective. Indeed, one would expect negative old items to have been encoded in memory, because they had been seen on many occasions during training. Thus, their memory trace should have been evoked when these old items were presented once again in the test phase. This is not the case for positive match items, because they were first presented at test and could at best evoke the imperfectly matching trace of the corresponding positive old item. As a consequence, on the basis of exemplar memory, one would expect larger response times for positive match than for negative old items.\(^2\)

The fact that the results were in the opposite direction can easily be explained on the basis of rule application. Note that categorization of four positive items (Items 1, 8, 9, and 15 in Table 1) requires only two feature tests, irrespective of the order in which the feature tests were performed, because these items either possess or lack all three attributes specified in the rule. In other words, they are the prototypes of the Tremblay or Beaulieu categories. The remaining positive items and all the negative items required either two or three feature tests, depending on the order of feature testing. Therefore, there was a clear advantage for positive items.

Recognition Test

Global performance was close to random (40% correct\(^3\)). For old items, there were 78% (SD = 23%) “old” responses; for match items, there were 70% (SD = 25%); and for new items, there were 74% (SD = 42%). We conducted signal detection analyses (Green & Swets, 1966) to evaluate each participant’s ability to discriminate the old items from the match and new items. It was found that participants could barely distinguish the old exemplars from the other two types of stimuli. The average d’ was .24 (SD = .63), and the average β was .84 (SD = .44). These results strongly suggest that the participants had not memorized the five feature combinations making up the training stimuli.

Discussion

Experiment 1 replicates the results obtained in Regehr and Brooks (1993; Experiments 1D and 2B). However, the nonrule attributes involved in our experiment were more salient than the rule attributes, and the participants were given 120 extra training trials. Despite these deliberate biases in favor of an exemplar effect, neither the error rates nor the response times obtained after 5 or 20 blocks of training showed a reliable negative match effect. The reliable difference between positive and negative items that we did obtain suggests instead that the application of the rule was guiding the participants’ categorizations.

The most likely explanation of our failure to obtain exemplar effects is that the participants did not have a good enough memory of the training exemplars. Let us examine briefly what to expect about categorization performance as a function of the number of attributes memorized. If the memory of the training exemplars encompasses all five attributes, then the presentation of a match item should evoke the trace of the nearest neighbor among old items according to Allen and Brooks’s (1991) line of argument. Facilitation or interference would also be expected if participants remembered only four of the stimulus attributes, provided that the set of four attributes memorized excludes back pattern. In this case, each match item becomes indistinguishable from the trace of its nearest old item, so that the negative match effect should be maximal. However, when another set of four attributes is remembered or when fewer than four attributes are remembered, traces of many different old stimuli can be equally similar with a given match item. As a consequence, there is no guarantee that presentation of a match item will evoke the trace of the old stimulus intended to produce facilitation or interference. In short, given Allen and Brooks’s stimulus structure, memory of the old exemplars needs to be fairly exhaustive in order for negative match effects to obtain when the stimuli are made of different combinations of interchangeable attribute values, as was the case here. Memory of all five attributes was also required for good recognition performance. The convergence of the categorization and recognition results suggests that exemplar memory was very partial, at best.

Another explanation may be proposed to account for the absence of a negative match effect in this experiment. In traditional category-learning tasks, participants do not need to focus on attributes with low cue validity to successfully classify the stimuli. Yet, for a negative match effect to occur in the rule application paradigm, the nondiagnostic attributes must receive attention, because they determine the necessary similarity between the old and the match stimuli for the effect to take place. Thus, it can be hypothesized that cue validity would generate incidental learning of nonrule attributes, which could in turn create exemplar effects in the rule application paradigm. We explored this possibility by conducting an experiment that was identical to Experiment 1, except that the nonrule attributes had as much cue validity as the rule attributes. The classification and recognition results were almost identical to those obtained in Experiment 1.

Experiment 2

In Experiment 2, we sought to test whether participants could memorize stimuli made of interchangeable attributes when explicitly instructed to do so, and whether this exemplar memory would influence the application of a categorization rule. As in Experiment 1, participants categorized stimuli using a perfectly predictive three-attribute rule. In addition, they were explicitly instructed to memorize all five attributes in association with a first name that was given to each creature. The goal of this name-learning task was to provide an opportunity for participants to attend to the nonrule attributes, in contrast with the categorization task that focuses the participants’ attention on the rule attributes. The experiment tested whether these instructions would foster the learning of the nonrule attributes and develop an exemplar memory of the stimuli that would be sufficient to generate negative match effects.

\(^2\) We are grateful to Philip Higham for bringing this aspect of the results to our attention, which is not to imply that Higham endorses the point that we make in the text, because his suggestions were of a methodological rather than theoretical nature.

\(^3\) We computed global means such as this one by weighing all items equally, which gives twice as much weight to new items compared to match items, because new items were two times more numerous.
Participants and Materials

Thirty-two students at the Université de Montréal participated in the study. All received Cdn$5 as compensation for their time. The stimuli and the classification rules were identical to those used in Experiment 1. However, in addition to classifying the creatures into two different families (the Tremblay and the Beaulieu), participants also had to learn a first name for each creature. The first names used were chosen to be short and fairly common in French. They were: Luc, Carl, Jean, Louis, Guy, Marc, Alex, and David. These names were partially counterbalanced with regard to their association with particular stimuli.

Procedure

All aspects of the procedure were identical to those used in Experiment 1, with two exceptions. First, the feedback for any given stimulus included both a family name (i.e., the category name) and a first name. The participants were instructed to take the 2-s period during which the creatures remained on the screen to learn the first name and to memorize the attributes of each individual creature. Second, after the recognition test, the participants were given a pencil-and-paper test to evaluate how well they had learned the creatures’ first names. For this identification test, each of the eight training stimuli was printed in color in one of eight numbered rectangles on a sheet of white paper. The participants were told to report the first name of each creature in writing. They could use each name only once.

Results

Classification Task

The analyses performed were the same as in Experiment 1. No participant was excluded from the analyses as a result of filtering out erroneous and excessively long trials.

Error rates. The error data pertaining to item type, item value, and number of training blocks are presented in the top portion of Figure 3. Error rates were larger than in Experiment 1 but were still quite low (all below 10%). The error rates were more pronounced for negative match (5.1%) than for negative old items (2.0%), and this difference was larger than that between positive match (2.0%) and positive old items (2.3%), resulting in a significant Item Type × Value interaction, F(1, 31) = 4.77, MSE = 0.00, p < .037. Figure 3 shows a tendency for the negative match effect to be more pronounced after 5 blocks of training than after 20 blocks, but the Item Type × Item Value × Number of Blocks interaction was not significant F(1, 31) = 1.26, MSE = 0.01, p = .27. Thus, instructions to memorize and associate the creatures’ attributes to a name succeeded in generating a conflict between exemplar memory and the application of the rule.

Response times. The response time data are presented in the bottom portion of Figure 3. The average difference between negative match and negative old items was in the right direction and was larger than the difference between positive match and positive old items. However, neither the Item Type × Value interaction, F(1, 31) = 1.44, MSE = 36.30, p = .24, nor the Item Type × Item Value × Number of Blocks interaction, F(1, 31) = .13, MSE = 24.28, p = .72, was significant. Once again, negative items (1,351 ms) were categorized more slowly than positive items (1,240 ms), F(1, 31) = 7.25, MSE = 110,36, p = .02.

Recognition and Naming Tests

Overall recognition accuracy was still at the 50% level. Participants did not succeed in recognizing the eight training stimuli even after studying each 20 times. For old items, there were 79% (SD = 41%) of “old” responses; for match items, there were 51% (SD = 50%); and for new items, there were 63% (SD = 48%). The average d’ was .77 (SD = .95), and the average β was .89 (SD = 1.27).

Participants’ ability to identify the eight training phase items by their first names was rated on a score of 0 (no good answers) to 8 (perfect score). The average was 5.34 (66%) with a standard deviation of 2.46. Eleven participants (35%) perfectly identified the eight stimuli, 5 (15%) identified six or seven of eight stimuli correctly, and 16 (50%) failed to correctly identify more than five stimuli. Although identification of the training stimuli could be achieved using only the rule attributes, it could also be performed on the basis of attribute combinations involving nonrule attributes. In the latter case, the scores on the naming task should be positively correlated with recognition performance (because the recognition test did require knowledge of some nonrule attributes in combination with some rule attributes). Over participants, the correlation between the naming scores and the d’ (old vs. other items) was significant, r(31) = .63, p < .001.

Discussion

The results of Experiment 2 are consistent with Brooks and his colleagues’ (Allen & Brooks, 1991; Regehr & Brooks, 1993) position that nonrule attributes can interfere with rule-based classification, even when such attributes are not related to category membership. However, it is one thing to claim that attributes, which receive little attention, influence categorization, and it is quite another to say that an attribute, which receives much attention, influences categorization. We were able to obtain an exemplar effect with stimuli made of interchangeable attributes only when participants were given intentional learning instructions, but the recognition test clearly showed that participants did not learn the five-attribute combinations making up the stimuli.

Nevertheless, the exemplar effect obtained in Experiment 2 was still modest. First, the negative match effect failed to generalize to response times. Second, even among error rates, the 9% average obtained with negative match items is quite low compared with the 45% obtained in Allen and Brooks’s (1991) Experiment 1, though it is quite comparable with the error rates obtained in some of Regehr and Brooks’s (1993) experiments. In Experiment 3, we investigated possible explanations for this discrepancy.

Experiment 3

Allen and Brooks’s (1991) original experiment differed from our Experiments 1 and 2 in three respects: the use of globally individuated stimuli, the use of backgrounds, and the presentation order of the stimuli at test. Experiment 4 deals with the issue of global individuation, and the remaining two issues are addressed here. Allen and Brooks wanted their task to have a certain degree of ecological validity. Hence, the creatures were shown on backgrounds depicting different living environments. Like the two
other nonrule attributes, the backgrounds were not diagnostic of category membership (see Table 1). However, participants in Allen and Brooks’s experiment were given intentional learning instructions concerning the backgrounds. On each trial of the training phase, participants had to classify the stimuli as builders or diggers on presentation of a first slide. Then, they had to “remember how the animal built or dug” (p. 6) while viewing two other slides. In the test phase, participants were asked to use background information to classify the stimuli: “. . . the first slide showed only the background on which the upcoming test item would be displayed. The subjects were simply to look at this background and indicate when they were ready for the second slide. The second slide showed the same background with a pair of animals on it . . . . They were also told that they might be able to use the first slide to anticipate which items were most likely to appear on the background . . . ” (p. 6).

Therefore, backgrounds had a special status compared with other nonrule attributes. The results of our Experiment 2 show that nondiagnostic attributes can be learned, albeit partially, when attention is drawn to them during training. Allen and Brooks’s (1991) instructions went beyond ours in that they encouraged participants to also use the information provided by the background during the test phase. This led us to hypothesize that the focus on the backgrounds may have been the key factor in causing negative match effects. To test this hypothesis, in Experiment 3, we presented stimuli composed of interchangeable attributes (as was the case in Experiments 1 and 2) on backgrounds. If negative match effects are found, the results will support the conclusion of

Figure 3. Mean error rates (A) and response times (B) obtained after 5 (open bars) and 20 (filled bars) blocks of training in Experiment 2. Error bars represent standard error of the mean.
Experiment 2 that the effects originate in nonrule attributes to which attention is paid.

Finally, Allen and Brooks’s (1991) Experiment 1 differed from our Experiments 1 and 2 (and from those of Regehr and Brooks, 1993) in that a nonrandom presentation order was used for the stimuli on the test phase. This methodological difference could therefore also have contributed to the differences in results across studies. Experiment 3 suggests that this is not the case, so we do not delve into the issue. Nonetheless, Experiment 3 involved two conditions. In one condition, backgrounds were introduced, and instructions were given to attract attention to them. In this background condition, the presentation order of the stimuli at test was random. We also used a presentation order + background condition in which the presentation order of the stimuli at test was as in Allen and Brooks’s original study.

Method

Participants and Materials

Sixty-four students at the Université de Montréal participated in the study. Each received Cdn$5 as compensation for their time. The stimuli were similar to those used in Experiment 1. They were built from five binary attributes: tail type (cane-shaped or stair-shaped), back pattern (stripes or spots), head shape (parabolic or oval), body type (oval or parallelogram), and number of legs (two or four). The first three attributes were part of the rule, and the last two attributes were not. All creatures were gray and were presented on one of four colored backgrounds (blue, white, yellow, or green). The backgrounds were nondiagnostic (see Table 1). The classification rules given to participants were also identical to those used previously.

Procedure

Participants were randomly assigned to one of two conditions: the background condition or the presentation order + background condition. Participants in the background condition received the training phase instructions given in Experiment 1. However, they were told to take the period of time during which the stimuli remained on display after the categorical decision had been made to notice the colored backgrounds and to relate them to the creatures with which they appeared. These instructions were parallel those given by Allen and Brooks (1991) in their Experiment 1 (1991, p. 6). The background condition included the five phases described for Experiment 1. In order, they were: 5 blocks of training with feedback, a test phase without feedback, 15 more blocks of training, another test phase, and an explicit recognition task.

The main differences between the procedure used for this experiment and that used in Experiment 1 were in the test phases. In addition to the instruction given in Experiment 1, participants were told that the colored backgrounds would appear before the creatures and that they could be used to anticipate the upcoming creature. These instructions also parallel those given in Allen and Brooks’s Experiment 1 (1991, p. 6). The number of trials used in the test phases of Experiment 3 also differed from those in Experiment 1. In this experiment, the old items were presented four times, and the match items were presented only once for a total of 40 trials. This number of trials is the same as used in Allen and Brooks’s original experiment. However, in the background condition of Experiment 3, the stimuli were presented in random order.

The procedure used in the presentation order + background condition was identical to that of the background condition except for the presentation order used during the test phases. The items were presented in a prespecified order identical to that of Allen and Brooks: positive old items (once), positive match items (once), negative old items (four times), positive old items (three times), and negative match items (once; see Allen & Brooks, 1991, p. 6, for a justification of this presentation order). Presentation order was random within each stimulus grouping.

Results

Classification Task

Analyses involved a $2 \times 2 \times 2 \times 2$ design with three within-subjects factors: item type (old vs. match), item value (positive vs. negative), and number of blocks (5 vs. 20); and one between-subjects factor: test condition (background vs. presentation order + background). Six participants were dropped because of incomplete or damaged data files because of computer problems. Filtering out the errors created empty cells for 17 participants. These participants were dropped from the response time analyses.

Error rates. The error data obtained after 5 and 20 blocks of training are presented in the top portion of Figure 4. The data shown are averaged over both the background and the presentation order + background conditions, because test condition did not produce a main effect, $F(1, 56) = 2.5, \text{MSE} = 0.18, p = .12$, and did not interact with any other factor (all $F$s < .9). Introducing the backgrounds caused a large increase in the error rates for negative match items. The error rates for negative match items (33%) were much larger than error rates for negative old items (5%), $F(1, 56) = 27.5, \text{MSE} = 4.35, p < .001$, although they were fairly small and constant for positive match (3%) and positive old items (5%), $F(1, 56) = 2.2, \text{MSE} = 0.01, p = .14$. As a consequence, the Item Type $\times$ Value interaction was highly significant, $F(1, 56) = 30.5, \text{MSE} = 2.4, p < .001$, but none of the higher level interactions involving these factors were (all $F$s < 1.5).

Response times. The mean response times obtained after 5 and 20 blocks of training are presented in the bottom portion of Figure 4. The data are averaged over both the background and the presentation order + background conditions. Once more, this factor failed to produce a main effect, $F(1, 39) = .36, \text{MSE} = 555,835, p = .55$, or to interact with any other factor (all $F$s < .57). In the condition most similar to that of Allen and Brooks’s (1991) original experiment, that is, after 5 blocks of training, the difference between negative match and negative old items was 323 ms. The corresponding difference in Allen and Brooks’s Experiment 1 was approximately 360 ms. The difference between negative match and negative old items appears smaller after 20 blocks than it does after 5 blocks, but number of blocks failed to interact with the other factors in the analysis. The difference between negative match and negative old items was nonetheless larger than the difference between positive match and positive old items in all subconditions of the experiments. This yielded a trend on the Item Type $\times$ Value interaction, $F(1, 39) = 3.3, \text{MSE} = 1,412,569, p = .077$. This interaction became significant when a log transform was applied to the data, $F(1, 39) = 7.8, \text{MSE} = 1.28, p = .008$, as was done by Allen and Brooks. The decomposition of this interaction showed that negative match items (1,181 ms) generated significantly longer response times than negative old items (924 ms), $F(1, 39) = 6.3, \text{MSE} = 2.4, p = .02$, whereas no such difference was present between positive match (817 ms) and positive old items (831 ms), $F(1, 39) = .5, \text{MSE} = 0.6, p = .83$. None of the higher order interactions achieved statistical significance in the analysis on log-transformed data.
Recognition Test

The data of 5 participants had to be excluded from analysis because of computer problems during the recognition test. For the remaining participants, recognition performance did not differ across the two test conditions. Participants had a success rate of 52% overall. For old items, there were 75% (SD = 19%) “old” responses; for match items, there were 60% (SD = 20%); and for new items, there were 53% (SD = 45%). The signal detection analyses once more showed that the participants’ ability to discriminate the old from the other stimuli was limited. The $d'$ averaged .69 (SD = .89), and the average $\beta$ was .78 (SD = .40).

Discussion

Although the results of Experiment 3 were similar to those originally obtained by Allen and Brooks (1991, Experiment 1), they warrant very different conclusions. First, the fact that error rates and response times both exhibited reliable negative match effects shows that stimuli do not have to be made of idiosyncratic attributes for exemplar memory to influence performance in rule-driven categorization. The error rates for negative match items soared from less than 10% in Experiments 1 and 2 to over 30% in Experiment 3. The negative match effect therefore appears to have resulted from the introduction of backgrounds, on which attention was explicitly brought to bear during the training and test phases.
To investigate whether the results obtained can be accounted for on the basis of the backgrounds, in conjunction with the rule attributes, we fit the version of the context model described by Smith and Minda (1998; see also Medin & Shaffer, 1978; Nosofsky, 1986) to the error rates obtained after 5 and 20 blocks. Each matching attribute made a contribution of 0.0 to the distance between stimuli, whereas each mismatching attribute contributed a distance that was equal to its attentional weight. A city-block metric was used to compute overall distance between stimuli, because the attributes were separable (Garner, 1974). No guessing parameter was used because the task was rule-based rather than inductive. Three versions of the model were fit using a maximum-likelihood criterion. The first one included the three rule attributes only. In this version, we fit three free parameters: the sensitivity parameter and two attentional weights (the third attentional weight was the difference between 1.0 and the sum of the other two attentional parameters). Another version included the rule attributes and the background (four free parameters), and the last version included all attributes (rule and nonrule) and the background (six free parameters).4

The results concerning performance obtained after 5 training blocks are shown in Table 2. Analysis of performance obtained after 20 blocks yielded similar results and led to the same conclusions. As can be seen in Table 2, the fit of the model based on only the rule attributes was poor (−ln L = 91.359, SSE = .1666, η² = .9368), showing that these three attributes alone do not suffice in accounting for the data. The model involving the rule attributes and the background (−ln L = 49.244, SSE = .0265, η² = .9899) fared significantly better, χ² (1, N = 2440) = 84.23, p < .001,5 suggesting that the background contributed to the effects obtained. The last model, comprising all attributes and the background (−ln L = 48.061, SSE = .0203, η² = .9927), did not fit significantly better than the preceding one, χ² (2, N = 2440) = 2.366, p > .10, suggesting that one can indeed account for the pattern of error rates without resorting to the remaining nonrule attributes.

How could the rule attributes and backgrounds be sufficient to yield a negative match effect? Although the backgrounds were nondiagnostic, the association of a given background with a given rule attribute often had cue validity. For instance, the two training items appearing on Background 1 (Items 2 and 6 in Table 1) could be distinguished and reliably categorized on the basis of head shape. The same was true of the two exemplars shown on Background 2 (Items 3 and 5 in Table 1). The two training exemplars shown on Backgrounds 3 (Items 4 and 8) could be reliably categorized on the basis of either tail type or back pattern; the same is true of the training items shown on Background 4 (Items 1 and 7). Because most of these background-attribute associations were preserved among transfer items (except those involving back pattern), they still pointed to the correct category name for positive match items. However, most of these associations now pointed to the wrong category name for negative match items. This led to the greater possibility of interference with the application of the rule.

This interpretation of the categorization results is consistent with the recognition data. An examination of Table 1 shows that memory for the three rule attributes in conjunction with the background would have led to perfect recognition of the old items and not to the random performances that were observed. The memory traces that included at least one rule attribute and the background were sufficient to yield negative match effects but insufficient to allow for successful recognition.

### Experiment 4

Experiments 1–3 suggest that the amount of attention given to nonrule attributes dictates their influence on rule-driven categorization. As a result, we can infer that the exemplar effect obtained in the rule application paradigm should not be related to a stimulus’ global form or gestalt. Yet, Regehr and Brooks (1993) believed that stimuli had to be globally individuated for negative match effects to take place. In Experiment 4, we investigated the origin of exemplar effects with stimuli made of idiosyncratic attributes.

The experiments that produced the largest negative match effects in Regehr and Brooks’s (1993) study involved stimuli that were individuated at both the attribute and global levels (Experiments 1D and 3A). Unfortunately, these experiments are ambiguous with respect to the issue of the global versus local origin of the exemplar effects in rule-based categorization. Indeed, when the attributes have the same perceptual implementation in many stimuli, as was the case in our Experiments 1–3 (see Figure 1), four of five attributes need to be taken into account in order for the match items to evoke their corresponding old items. However, when the attributes have a unique perceptual implementation in each pair of old and match items (see Training Set A and Transfer Set A in

---

**Table 2**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model</th>
<th>Rule attributes only</th>
<th>Rule and background attributes</th>
<th>Rule, nonrule, and background attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>7.858</td>
<td>8.344</td>
<td>8.102</td>
<td></td>
</tr>
<tr>
<td>wTail type</td>
<td>.373</td>
<td>.318</td>
<td>.337</td>
<td></td>
</tr>
<tr>
<td>wBack pattern</td>
<td>.297</td>
<td>.218</td>
<td>.232</td>
<td></td>
</tr>
<tr>
<td>wHead shape</td>
<td>.330</td>
<td>.329</td>
<td>.292</td>
<td></td>
</tr>
<tr>
<td>wTail type</td>
<td>.569</td>
<td>.069</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>wNumber of legs</td>
<td>.135</td>
<td>.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wBackground</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fit</td>
<td>−ln L</td>
<td>91.359</td>
<td>49.244</td>
<td>48.061</td>
</tr>
<tr>
<td>% variance accounted for</td>
<td>93.683</td>
<td>98.989</td>
<td>99.266</td>
<td></td>
</tr>
</tbody>
</table>

Note. c = generalized sensitivity parameter; wₘ = attentional weight given to dimension m; SSE = sum of squared deviations between observed and predicted classification probabilities. −ln L = negative value of log likelihood.

---

4 We fit the models to the group data because there were too few observations per items (only one per match item) to fit the models to the performance of every individual participant.

5 We used the likelihood-ratio testing method (Wickens, 1982) to test whether a model with more parameters fit significantly better than a nested model with fewer parameters.
Regehr and Brooks (1993) were aware of this possibility and designed a follow-up Experiment (3B) to generate evidence against it. The new experiment used the same design as the previous experiments except that the two nonrule attributes were different for transfer items (see Training Set B and Transfer Set A in Figure 5). Regehr and Brooks hypothesized that if the preservation of the global aspect of the stimulus was necessary for negative match effects to occur, then modifying the nonrule attributes would eliminate them. Regehr and Brooks obtained this result and argued that the negative match effects resulted from a conflict between the application of the rule and the global similarity of the old and corresponding match items. This conclusion suggests that the exemplar effects obtained with highly individualized stimuli are different from the effects we obtained with stimuli made of interchangeable features. As a consequence, we decided to duplicate and extend Regehr and Brooks’s Experiments 3A and 3B.

Regehr and Brooks used different training sets in Experiments 3A and 3B and did not control for possible differences in learning between the two. Hence, the differences in results between Experiments 3A and 3B may have been unrelated to Brooks and his colleagues’ hypothesis, stemming instead from dissimilar training experience. This shortcoming is not trivial from a similarity perspective. Changing nonrule attributes produces creatures with different global looks. Regehr and Brooks supposed that the individuality of the creatures played an important role in generating the conflict between old and match items, but creature salience, attractiveness, and ease of memorization were not controlled when using different sets of training exemplars. In Experiment 4, the two training sets illustrated in Figure 5 were factorially combined with the two transfer sets.

We also extended Regehr and Brooks’s work by using two different designs for the data analyses in Experiment 4. The first design was used in Regehr and Brooks’s (1993) study. It merged the data for all old items, positive and negative, and compared them with those obtained with positive and negative match items separately. We believe that this design is problematic because it can suggest that a negative match effect exists even when the difference between negative match and negative old items is similar to the difference between positive match and positive old items. Hence, it cannot support unambiguously the interpretation that exemplar memory influenced rule use. The second design was used in Allen and Brooks’s (1991) study and in our Experiments 1–3. It involves analyzing the difference between negative match and negative old items versus the difference between positive match and positive old items. When an interaction is found, as was the case in Experiment 3, the data show ostensibly that rule application is more error-prone and slower for negative match items than it is for positive match items, controlling for possible differences in performance among the corresponding training items. As a consequence, this design makes it difficult to support alternative interpretations of the data.

Method

Participants and Materials

Sixty-four students at the Université de Montréal were randomly assigned to one of four conditions resulting from crossing the two training sets with the two transfer sets in Figure 5. Each participant received Cdn$3 as compensation for her or his time.

The stimuli were drawings of fictional animals built from five binary attributes: number of legs (six or two), spots (present or absent), body type (angular or curved), neck length (long or short), and tail length (long or short). Each attribute value was implemented idiosyncratically for all training exemplars. The first three attributes specified category membership. We used the same four rules as Regehr and Brooks (1993) within each condition. These rules were: (a) six legs, angular body, and spots present; (b) two legs, angular body, and spots present; (c) six legs, round body, and spots absent; and (d) two legs, round body, and spots absent. The task required that a creature have two or three of the specified attribute values in order to be classified in a given category (i.e., as a Maurice). Otherwise, it had to be classified as an Henri. The last two attributes, neck and tail length, were not part of the rule. We constructed the match items by changing the spot attribute of the training items. The abstract structure of the stimuli was the same as in Experiments 1–3 (see Table 1).

Old items were selected from Training Set A or B, and match items were selected from Transfer Set A or B, depending on the group of participants.
In conditions in which the nonrule attributes remained unchanged, the old and corresponding match items were identical except for the spot attribute. This created the four familiar item types: positive old, negative old, positive match, and negative match. However, in conditions in which the nonrule attributes were modified, only the attributes body type and number of legs were identical for both old and corresponding match items.

Items in Figure 5 that were not seen by the participants during the training and test phases of the experiment were used as new items in an explicit recognition memory task. For example, participants who were exposed to Training Set A and Transfer Set A were presented with creatures from the Training and Transfer Sets B as new items in the memory test. Although the specific creatures used in the classification task changed over groups of participants, there were always 16 new items remaining for the recognition test.

**Procedure**

In Experiment 4, we used the five experimental phases that were used in Experiment 1. To begin, the participants were given the categorization rule and instructed to classify the creatures accordingly. They were given 40 trials divided in 5 blocks. However, unlike in Experiment 1, the stimuli disappeared as soon as the participants’ answers were recorded. The feedback was displayed alone on a dark background. In the first test phase, the eight old items, the four positive match items, and four negative match items were presented once each in random order. This phase of the experiment proceeded like the training phase except that no feedback was given. At this point, Regehr and Brooks’s (1993) procedure (Experiments 3A and 3B) had been duplicated. In the third and fourth phases of the experiment, the participants were given an additional 15 blocks (120 trials) of training followed by a second test, identical to the first one.

The last phase was an explicit recognition test. Participants were shown the 8 old items, the 8 match items, and the 16 new items in a random order. Their task was to determine whether a particular item had been seen in one of the first four phases of the experiment or whether it was new. Contrary to the recognition test used in Experiments 1–3, participants had to discriminate old and match items, on the one hand, from new items, on the other. This was done to test whether participants could remember some of the nonrule attributes seen during the categorization phases. If this is the case, the recognition task should be easier for participants who see only a subset of nonrule attributes in the categorization task than for those who see all of them, because the former only need to recognize one attribute that was never seen before in order to identify a stimulus as new, whereas the latter must remember combinations of rule and nonrule attributes to accomplish the recognition task. Participants responded by selecting the appropriate key on the keyboard (1 or the space bar). No feedback concerning response accuracy was given.

**Results**

**Classification Task**

To facilitate comparison with Regehr and Brook’s (1993) results, we present performance obtained after 5 blocks of training separately from performance obtained after 20 blocks.

**Five blocks.** The error rates and response time results were first analyzed following Regehr and Brook’s (1993) method, except that all training and transfer set combinations were included in the design. The $2 \times 2 \times (3)$ ANOVAs included one within-subjects factor: item (old [both positive and negative] vs. positive match vs. negative match) and two between-subjects factors: training set assignment (A vs. B) and test condition (same nonrule attributes in the training and transfer sets vs. different nonrule attributes in the training and transfer sets). To be concrete, this last between-subjects factor pitted participants having received Training Set A followed by Transfer Set A or Training Set B followed by Transfer Set B against those having received A followed by B and B followed by A. Error trials and excessively long trials were eliminated from the response time analyses. This created an empty cell for 1 participant whose data were dropped from the response time analyses. As we describe shortly, training set was never found to have any significant effect on either error rates or response times, and it was not involved in any significant interaction. So we do not present the results obtained with the two training sets or discuss them separately.

Table 3 shows the mean error rates and response times obtained in the same versus different nonrule attribute conditions. Over both conditions, mean error rates were 9% for old items, 12% for positive match items, and 22% for negative match items, $F(2, 120) = 14.7, MSE = 0.3148, p < .001$. The response times, in the same order, were 1,947 ms, 2,048 ms, and 2,459 ms, $F(2, 120) = 14.9, MSE = 4,702.276, p < .001$. Although item type was significant for both error rates and response times, it did not interact singly or jointly with the factors training set or test condition in either the error rate analysis (all $Fs < 1.4$) or the response time analysis (all $Fs < 1.5$). Hence, both variables showed a disadvantage for negative match items. Regehr and Brooks (1993; Experiments 1D, 3A, and 3B) took analogous results as evidence for a conflict between similarity-based and rule-based categorization.

The data were submitted to a second type of analysis, one involving the previously used Item Type (old vs. match) $\times$ Item Value (positive vs. negative) design. Apart from the two within-subjects factors mentioned, the analysis included the same two between-subjects factors as above: training set (A vs. B) and test condition (same vs. different nonrule attributes).

**Table 3**

*Mean Error Rates (in Percentages) and Response Times (in ms) Obtained After Five Blocks of Training in the Same vs. Different Nonrule Testing Conditions in Experiment 4*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Same nonrule attributes</th>
<th>Different nonrule attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Error rates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive old</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Negative old</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Old (average)</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Positive match</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Negative match</td>
<td>18</td>
<td>27</td>
</tr>
<tr>
<td><strong>Response times</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive old</td>
<td>1,936</td>
<td>1,626</td>
</tr>
<tr>
<td>Negative old</td>
<td>2,237</td>
<td>1,773</td>
</tr>
<tr>
<td>Old (average)</td>
<td>2,087</td>
<td>1,700</td>
</tr>
<tr>
<td>Positive match</td>
<td>2,130</td>
<td>1,937</td>
</tr>
<tr>
<td>Negative match</td>
<td>2,665</td>
<td>2,186</td>
</tr>
</tbody>
</table>

*Note.* Averaged data for old items are presented to facilitate comparisons with Regehr and Brook’s (1993) Experiments 3A and 3B.
The error rate data did not reveal an Item Type \( \times \) Value interaction, \( F(1, 60) = .381, MSE = 0.006, p = .539 \). The overall difference between negative match (22\%) and negative old items (13\%) was 9\%, and the difference between positive match (12\%) and positive old (5\%) was 7\%. These factors were not involved in any interaction involving training set (both \( F < .4 \)), but there was a very slight trend for item type, item value and test condition to interact, \( F(1, 60) = 2.579, MSE = 0.04, p = .114 \).

The response time data were in the direction predicted by Regehr and Brooks (1993). Globally, the difference between negative match (2,426 ms) and negative old items (2,005 ms) was 421 ms, and the difference between positive match (2,034 ms) and positive old items (1,781 ms) was 253 ms. However, the item type and item value factors were not found to interact significantly with each other, \( F(1, 59) = 2.159, MSE = 445,510, p = .147 \), and they did not interact with the factors training set, test condition, or both (all \( F < 1.2 \)). Hence, the analyses based on the Item Type (old vs. match) \( \times \) Item Value (positive vs. negative) design show that none of the interactions were reliable.

Twenty blocks. We analyzed the data obtained after 20 blocks of training using only the \( 2 \times 2 \times (2) \times (2) \) design, which included the within-subjects factors training set and test condition, and the two between-subjects factors item type and item value. Eliminating error trials from the response time data created empty cells for 3 participants. Their data were dropped from the response time analyses. The mean error rates and response times are shown in Figure 6.

![Figure 6](image_url)
For error rates, the Item Type \( \times \) Item Value interaction was significant, \( F(1, 60) = 9.8, MSE = 0.21, p = .003 \). A decomposition of this interaction revealed that the 20\% difference between negative match (26\%) and negative old items (6\%) was also significant, \( F(4, 60) = 10.4, MSE = 0.03, p < .001 \). However, Item Type \( \times \) Value did not interact with training set or test condition either singly or jointly (all Fs < 1.4). Hence, the prolonged learning period produced a conclusive negative match effect, but the effect was unrelated to whether the nonrule attributes were the same or different at the time of transfer. The same pattern of results was obtained for response times. The Item Type \( \times \) Value interaction was significant, \( F(1, 57) = 16, MSE = 2,769,969, p < .001 \), and the decomposition of this interaction showed that the 904-ms difference between negative match (1,893 ms) and negative old items (989 ms) was also significant, \( F(4, 57) = 15.1, MSE = 6,339,099, p < .001 \). Once more, there were no significant higher order interactions involving item type and item value (all Fs < 1.1).

**Recognition Test**

Recognition was 91\% correct when the new stimuli had novel nonrule attributes. For old items, there were 94\% (SD = 9\%) correct "other" responses; for match items, there were 87\% (SD = 13\%); and for new items, there were only 9\% (SD = 11\%) incorrect "other" responses. The average \( d' \) was 3.05 (SD = 95), and the average \( \beta \) was 2.35 (SD = 3.23).

In contrast, the success rate was only 64\% when the new stimuli were made of nonrule attributes encountered during training and test phases. For old items, there were 90\% (SD = 12\%) "other" responses; for match items, there were 58\% (SD = 20\%); and for new items, there were 47\% (SD = 17\%). The average \( d' \) was .78 (SD = .60), and the average \( \beta \) was .85 (SD = .28). A t test on \( d' \) confirmed that it was much easier to distinguish old and match items from new items when the former were made of novel attributes than when they were made of novel combinations of previously seen attributes, \( t(59) = 11.1, p > .001 \).

**Discussion**

In Experiment 4, we succeeded in obtaining reliable negative match effects under incidental learning conditions in the rule application paradigm. However, this required more extensive training than Regehr and Brooks’s (1993) results led us to expect. More important, the effects obtained did not depend critically on maintaining the same nonrule attributes throughout the categorization task, because negative match effects were also present when the nonrule attributes were different for transfer items. Changing the nonrule attributes at transfer did modify the global form of the stimuli, but this did not weaken the negative match effects observed. The results of Experiment 4, therefore, support the conclusions reached in Experiments 1–3. Only attributes that received the participants’ attention influenced rule application in this paradigm. The preceding conclusions would hardly be surprising if participants had failed to notice the nonrules attributes during training. However, this does not seem to be the case. In the recognition test, the participants who best discriminated the old from the new items (same nonrule attributes condition) could do so on the basis of the novel nonrule attributes. This result suggests that the nonrule attributes were noticed during training. The poor recognition performance of the other condition shows that novel combinations of attributes were not so easily noticed, which is consistent with the suggestion that the negative match effects obtained in the categorization task stemmed from the rule attributes.

**General Discussion**

We have presented four experiments in which we reexamined the influence of exemplar memory on the application of a categorization rule. In Experiment 1, we failed to elicit a negative match effect using stimuli made of interchangeable attributes, prolonging the training period, and making the nonrule attributes more salient. In Experiment 2, we explicitly instructed participants to learn the stimuli during the categorization task. This procedure did elicit a small negative match effect in error rates, and yet the recognition data showed that the participants’ memory for the stimuli was negligible. In Experiment 3, we showed that a nonrule attribute receiving systematic attention could create strong negative match effects in both error rates and response times, although the attributes were interchangeable. Finally, Experiment 4 showed that the rule-based categorization of stimuli composed of idiosyncratic attributes could also lead to negative match effects but that the effects are related to the rule attributes.

Allen and Brooks (1991) had suggested that memory for exemplars develops incidentally, rapidly, and comprehensively, without regard for the diagnostic value of the individual attributes comprising the stimuli, and that this memory exerts a strong influence on the application of a categorization rule, even when it is well known and practiced. Regehr and Brooks (1993) added that idiosyncratic attributes and globally individuated stimuli were necessary for negative transfer effects to occur. Our results did not support either of these views.

We did obtain effects of exemplar memory on rule-driven categorization. However, the phenomenon seemed to depend on selective attention (Kruschke, in press) guided by the application of the rule or the experimental instructions. The phenomenon occurred when the resulting, incomplete memory traces could support correct category decisions during training, while occasionally leading to incorrect test decisions (Experiments 2–4). Moreover, the strength of the phenomenon appeared to be inversely related to the number of attributes in these memory traces. The phenomenon was weaker when large combinations of attributes were needed to make memory-based categorical decisions (Experiment 2) than when categorization decisions could be based on the memory of only two attributes (the background and one other attribute, Experiment 3) or even one (due to its idiosyncratic nature, Experiment 4). The recognition data are perfectly in line with this account of the categorization results, because the participants could not recall the five-attribute combination of attributes that composed the stimuli even when the attributes were idiosyncratic.

To further support this position, we submit that our conclusions regarding negative match effects in the rule application paradigm seem to be more consistent with typical findings in the categorization literature. A variety of categorization experiments have shown that participants find it very difficult to learn complex
relationships between stimulus attributes and category membership.

First, studies that have compared participants’ ability to predict a specific attribute belonging to a given stimulus using the other attribute values and the category label as cues typically show poor results if they previously learned to classify the stimuli via induction (Anderson, Ross, & Chin-Parker, 2002; Chin-Parker & Ross, 2002; Yamauchi & Markman, 1998). However, performances are superior if participants are asked to predict attributes from other attributes and the category label during the training phase of these experiments (also see Lassaline & Murphy, 1996). These studies show that participants are unable to recall within-stimulus attribute relationships unless they have focused their attention directly on them.

Undirected sorting tasks offer convergent evidence for this last idea (Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987; Murphy, 2002). Participants typically exhibit a strong tendency to sort the stimuli according to a single attribute, although the categories are based on family resemblances. More important, they show very little evidence of having encoded family resemblances in a way that would allow them to categorize the stimuli on the basis of global similarity (Regehr & Brooks, 1995). When sorting stimuli with one-attribute rules, participants even typically fail to notice attributes that are perfectly correlated with the rule attribute (Giguère, Lacroix, and Larochelle, 2004). In addition, research that has focused directly on unsupervised category learning has shown that participants’ attention must be focused incidentally on the stimulus attributes in order for learning to take place (Love, 2003). Otherwise, categorization performance is poor.

Finally, our position is consistent with induction paradigm results (Allen & Brooks, 1991; Medin et al., 1982; Medin & Schaffer, 1978; Nosofsky, 1986). In these experiments, the participants have no knowledge about category membership, and they must learn to classify the stimuli from feedback. Research has shown that participants typically proceed by testing different one-attribute rules. When a rule is deemed sufficiently successful, the participants learn the stimuli that are exceptions (Nosofsky, Palmeri, & McKinley, 1994). Under these learning conditions, the participants may focus on diagnostic and nondiagnostic information, while they are testing different rules, before converging on the set of attributes that maximizes categorical decision. Thus, they encode stimulus information that can create an interaction between exemplar memory and the inferred rule stemming from both the rule or nonrule attributes.

Thus, in agreement with our experiments, categorization research shows that participants do not easily learn the interattribute relationships that enter in the composition of stimuli. When they do learn some of these relationships, evidence that the participants’ attention was brought to bear on them is always present. As a last remark, we add that this attentional requirement is in agreement with the episodic approach that Brooks has taken in recent work (Brooks & Hannah, 2002).

Accounting for These Findings

It is clear that our analyses have shown that the generalized context model (GCM; Nosofsky, 1986) can account quite accurately for categorization performance observed in Experiment 3, and it would most likely do as well for the results of our other experiments. One can assume that ALCOVE (Kruschke, 1992) would learn to categorize the stimuli in our experiments. However, the fact that the categorization rule operates successfully from the onset of the experiment and that it can interact with a partial attribute-based memory of the exemplars makes these models unattractive explanations for our results. In their defense, one must add that these models were not designed to simulate categorization data obtained when a categorization rule is available from task onset.

ATRIUM constitutes a more plausible alternative. This hybrid model includes a rule module that draws categorical boundaries in the psychological space and an exemplar-learning module implemented with ALCOVE. In a section of their general discussion in which they specifically addressed Allen and Brooks’s (1991) rule application paradigm, Erickson and Kruschke (1998, p.127) speculated that ATRIUM’s rule module would learn to classify all the stimuli accurately, whereas the exemplar module would learn to associate the stimuli with the category responses. Then, at transfer, they argued that the similarity between old and match items would lead ATRIUM to incorrectly classify some stimuli. The GCM simulations that we conducted in Experiment 3 support Erickson and Kruschke’s conjecture. However, as we have shown, the simulation would work with the exemplar module allocating little or no weight to the nonrule attributes.

Competition frameworks such as the competition between verbal and implicit systems model (COVIS; Ashby, Alfonso-Reese, Turken, & Waldron, 1998) and Logan’s model of automatization (Logan, 1988, 1992) can also accommodate the exemplar effects in the rule application paradigm. COVIS is a hybrid model. It has a procedural system that learns to associate category responses with different regions of psychological space to create optimal decision boundaries (see Ashby et al., 1998). This system competes with a verbal system to deliver the categorization responses. With practice, the competition usually favors the implicit system because it becomes more accurate, but the verbal system will still respond sporadically. In our experiments, the verbal system would have supported perfectly accurate categorizations.

However, accuracy is not the only justification for an implicit system such as that postulated in COVIS. Efficiency is another one. According to Logan (1988, 1992), it is assumed that the early stages of learning rely mostly on algorithm-based processing. People first have to follow the set of rules necessary to perform a given task. Gradually, as episodic memory of the stimuli in association with the appropriate responses develops, one can more easily perform the task by accessing the relevant memory traces. Thus, prolonged training usually favors exemplar-based processing (see also Smith & Minda, 1998). This is what we observed in Experiment 4 when categorization results were compared after 5 and 20 blocks of training. This competition framework is also consistent with our finding that exemplar effects were more pronounced when the number of attributes considered by the memory system (two in Experiment 3 and one in Experiment 4) was smaller than the number of attributes in the rule.

In summary, our results do not negate the influence of exemplar memory on rule-based mechanisms. However, exemplar learning does not seem to proceed as autonomously as Allen and Brooks (1991) and Regehr and Brooks (1993) initially suspected. Instead of encompassing the whole stimuli, exemplar learning appears to
be guided by and limited to the attributes mentioned in the instructions. In accordance, the exemplar effects obtained at transfer do not seem to originate in the stimulus gestalt but in the attributes that received attention during training and at the time of test. As such, our results are perfectly compatible with encoding specificity (Tulving, 1983; Tulving & Thompson, 1973) and transfer-appropriate processing (Morris, Bransford, & Franks, 1977). Also, from a larger ecological perspective, one would think that exemplar-based mechanisms are involved in categorization because they generally help rather than hinder processing. This would not be the case if we were endowed with a powerful similarity-based mechanism that learned irrelevant information without attention and that used this information to override conscious rule application. So, it makes much more sense that the attributes involved in similarity-based processing be limited to those that are relevant to the task at hand, which are generally the rule attributes in Allen and Brooks’s paradigm.

References


