COGNITION

Cognition 65 (1998) 167–196

Alternative strategies of categorization

Edward E. Smith*, Andrea L. Patalano, John Jonides

Department of Psychology, University of Michigan, 525 East University, Ann Arbor, MI 48109, USA

Abstract

Psychological studies of categorization often assume that all concepts are of the same general kind, and are operated on by the same kind of categorization process. In this paper, we argue against this unitary view, and for the existence of qualitatively different categorization processes. In particular, we focus on the distinction between categorizing an item by: (a) applying a category-defining rule to the item vs. (b) determining the similarity of that item to remembered exemplars of a category. We begin by characterizing rule application and similarity computations as strategies of categorization. Next, we review experimental studies that have used artificial categories and shown that differences in instructions or time pressure can lead to either rule-based categorization or similarity-based categorization. Then we consider studies that have used natural concepts and again demonstrated that categorization can be done by either rule application or similarity calculations. Lastly, we take up evidence from cognitive neuroscience relevant to the rule vs. similarity issue. There is some indirect evidence from brain-damaged patients for neurological differences between categorization based on rules vs. that based on similarity (with the former involving frontal regions, and the latter relying more on posterior areas). For more direct evidence, we present the results of a recent neuroimaging experiment, which indicates that different neural circuits are involved when people categorize items on the basis of a rule as compared with when they categorize the same items on the basis of similarity. © 1998 Elsevier Science B.V.

Keywords: Categorization; Similarity; Rule application; Neuroimaging

1. Introduction

Since the beginning of the experimental study of categorization in psychology (Hull, 1920), there has been a tendency to assume that all acts of categorization are accomplished by the same means. In the seminal studies of Bruner et al. (1956),

* Corresponding author. Tel.: +1 313 7640186; e-mail: eesmith@umich.edu

PII S0010-0277(97)00043-7

people were assumed to rely primarily on rules when categorizing novel items; in the current models of Estes (1994) and e.g. Nosofsky (1992), people are assumed to categorize new objects solely on the basis of their similarity to remembered exemplars of known categories; and in the discussions by Keil (1989) and Rips (1989), categorization inevitably comes down to applying a 'theory' of the category. The alternative to such unitary views, of course, is that there are multiple strategies of categorization. This issue of unitary vs. multiple strategies is of foundational importance. If there are multiple strategies or procedures but we act as though there is only one, then results from different situations will no doubt conflict, with few or no generalizations emerging and no true accumulation of research findings[1].

Though much of the research motivated by the various unitary views is of great importance, it is not difficult to devise counterexamples to the idea of a unitary view. One can generate pairs of extreme situations, such that just about everyone would agree that categorization is based on one strategy in one case and on a different procedure in the other case. Thus, if a person has to categorize two-digit numbers as odd or even, presumably all researchers would agree that the categorizer does it by applying the rule of 'Divisible by 2 or not' (Armstrong et al., 1983); but if that same person has to categorize novel people with respect to whether they are as friendly as the neighbors on their block, presumably most would agree that the categorizer now relies on memories of his or her neighbors (Kahneman and Miller, 1986). Note, however, an important aspect of this counterexample—the different putative categorization procedures are applied to different kinds of categories, where one category is part of a rule-based formal system and the other one is completely *ad hoc*. If we ask instead whether there are multiple categorization procedures that are routinely applied to the same categories, then the issue of unitary versus multiple strategies is very much alive. It is this more stringent question that is the subject of the present paper.

Another aspect of the issue of one or many procedures is whether the multiple procedures are qualitatively different from one another. Some have argued for multiple categorization strategies that are qualitatively similar. For example, Nosofsky et al. (1994) have proposed a rule-plus-exception model, in which the representations used by the two procedures—rules and stored exemplars—differ quantitatively rather than qualitatively (e.g. the representation of a simple rule specifies a single attribute value, whereas the representation of a remembered exemplar might specify two or three attribute values). The obvious alternative is that there are qualitatively different procedures, that is, procedures that contain different processes.

In this article we argue that people can apply multiple procedures of categorization to the same items, perhaps even simultaneously, and that the procedures are qualitatively different from one another. Our specific agenda for the remainder of the paper is as follows. In the next, or second section we select two categorization procedures for examination—applying a rule vs. determining similarity to remembered examples—and characterize each procedure in more detail. In the third sec-

[1]We use the terms 'strategy' and 'procedure' interchangeably. However, since 'strategy' suggests the notion of deliberative choice, we favor the term 'procedure' wherever this suggestion would be misleading.

tion we review evidence from cognitive studies that supports the claim that the two procedures of interest are applied to the very same categories. This review will accomplish two goals. Firstly, it will integrate findings from experiments that have used artificial materials with results from studies that have employed natural categories, showing striking convergence between two literatures that have heretofore been kept separate. Secondly, our review will reveal some of the processes that comprise the two procedures of interest, which will show that the procedures are indeed qualitatively different. In the fourth section we consider a different kind of evidence for a qualitative difference between the procedures of interest. Specifically, we will review results from neuropsychology and neuroimaging experiments which indicate that the procedures at issue are mediated by different neural structures. The fifth and final section summarizes our main points, and notes some related research.

2. Rule application and exemplar similarity

2.1. The general distinction

It is time to be more precise about possible categorization procedures. A review of the literature on concepts and categorization (Smith and Medin, 1981) suggests at least three distinct procedures. In deciding whether a test object belongs to a particular category, one may:

1. Determine whether the test object fits a rule that defines the category (the rule specifies the necessary and sufficient conditions for category membership);
2. Determine the similarity of the test object to remembered exemplars of the category; or
3. Determine the similarity of the test object to a prototype of the category.

Subsequent work (e.g. Murphy and Medin, 1985) led to the addition of another strategy to the list:

4. Determine whether the features of the test object are best explained by the ‘theory’ that underlies the category.

We will focus on just the first two of these classification procedures—which we will refer to as ‘rule application’ and ‘exemplar similarity’—because they are among the most widely discussed in the literature, and because they are sufficiently conceptually distinct that one may readily tell them apart in both cognitive and neuropsychological studies[2].

To illustrate paradigm cases of the procedures of interest, consider the situation in which a dermatologist must decide whether a particular skin lesion is an instance of Disease X. (The following examples are inspired by Brooks et al., 1991). Suppose that our dermatologist knows the additive rule that, ‘If the lesion has a sufficient

[2]The distinction between rule-application vs. exemplar-similarity mechanisms has also been raised in the category-learning literature (e.g. Shanks and St. John, 1994). In this paper, we focus on mechanisms guiding category use.

number of the following features—elliptical shape, bumpy texture, reddish-brown coloring, etc.—then Disease X is indicated’. If the dermatologist applies this rule in making her diagnosis (categorization), then presumably she will engage in the following sequence of processes:

1. Selectively attend to each critical attribute of the test object (e.g. the shape, texture and color of the lesion);
2. For each attended-to attribute, determine whether the perceptual information instantiates the value specified in the rule (e.g. ‘Is this color reddish-brown’?); and
3. Amalgamate the outcomes of Stage (2) so as to determine the final categorization.

This three-stage schematic model of rule application is compatible with numerous discussions of rule following (e.g. Kemler-Nelson, 1984; Smith and Sloman, 1994).

Categorization based on exemplar similarity is a very different matter. Consider, again, our dermatologist attempting to diagnose a particular lesion. In addition to knowing the above rule, if the dermatologist has also seen many patients she will likely have stored in memory numerous exemplars of various skin diseases. Consequently, she may note that the current lesion is very similar to stored exemplars of Disease X, and on this basis categorize the current lesion as an instance of X. Now the sequence of processes presumably include:

1. Retrieve stored exemplars (of various disease categories) that are similar to the test object; and
2. Select that category whose retrieved exemplars are on some measure most similar to the test object.

Note that if the exemplars retrieved in Stage (1) all belong to the same category, then the choice process of Stage (2) is trivial. But if the exemplars retrieved in Stage (1) point to different categories, Stage (2) might require an effortful selection process; for example, computing the similarity of the test object to each retrieved exemplar, combining these similarity scores over members of a category, and choosing the category with the highest similarity score. (Of course, a similarity process is likely the basis of Stage (1) retrieval as well, but presumably it is a more automatic and holistic process). This schematic description of exemplar-based categorization captures some of the key ideas behind the major exemplar-similarity models (e.g. Estes, 1994; Nosofsky, 1986; Medin and Schaffer, 1978).

2.2. Component distinctions

The preceding discussion suggests that the general contrast between rule application and exemplar similarity includes a number of component or correlated distinctions. Specifically, the two categorization procedures differ in the extent to which they involve: (a) analytic vs. holistic processing, (b) differential vs. equal weighting of attributes, (c) instantiation of abstract conditions vs. matching concrete information, (d) high vs. low loads on working memory, (e) serial vs. parallel processing, and (f) strategic vs. automatic processing. We briefly consider each of these component distinctions.

We noted that rule application involves selectively attending to the critical attributes of the test object (and perhaps inhibiting others). This selective-attention component makes rule-application an analytic procedure. In contrast, the retrieval of similar stored exemplars, which comprises the heart of the exemplar-similarity procedure, need not involve any selectivity, and in this sense is often referred to as a holistic process.

Because rule application involves attending to some attributes but not others, the procedure gives different weights to different attributes. In contrast, because exemplar similarity need not assume any selective attention, at least during the critical exemplar-retrieval stage, the procedure may give the same weight to all attributes. A distinction between differential and equal weighting of attributes is, therefore, a natural consequence of the analytic-holistic distinction.

In paradigm cases of rule application, the conditions specified in the rule are more abstract than the representation of the test object. Consequently, rule application typically requires that the categorizer determine whether the information in the test object instantiates the conditions of the rule. In many cases of exemplar similarity, though, the representations of both exemplar and test object are assumed to be at the same level of concreteness, and hence a matching process rather than an instantiation one is needed.

Working memory is often involved in rule application for one of two reasons. In some cases, the rule is sufficiently novel or complex that the categorizer needs to keep it active; in other cases, the rule has numerous conditions, and the categorizer must keep active the outcomes of prior condition tests while performing subsequent ones. Taken together, these reasons could lead to a substantial load on working memory. In contrast, the retrieval of exemplars from long-term memory may impose a relatively small load on working memory, particularly in situations in which only a single exemplar is retrieved.

Note that the distinctions just described are closely related to the three-stage process we used to characterize rule application (see Section 2.1). Being analytic and differentially weighting attributes characterizes the first or selective attention stage, instantiation of abstract-conditions captures the second or instantiation stage, and an involvement of working memory defines the third or amalgamation stage. All of this reinforces the point that rule application is a complex procedure that contains at least three major mechanisms.

The remaining two distinctions do not pick out component mechanisms of rule application or exemplar similarity, but rather characterize the operation of the components. Thus, while exemplar similarity may involve some serial processing (a retrieval process *followed* by a selection process), serial processing seems more pronounced in rule application. This is particularly so when the rule specifics multiple conditions, which may require multiple acts of selective attention and instantiation.

Lastly, paradigm cases of rule application usually involve strategic or controlled processing, whereas paradigm cases of exemplar similarity typically involve more automatic processing. Some aspects of this strategic vs. automatic distinction are captured by previous distinctions, as more strategic processing is more likely to differentially weight information, require working memory, and involve serial processing. Still, there are other aspects of the present distinction that are novel; for example, rule application, being more strategic, is easier to verbalize.

While the preceding distinctions are useful in distinguishing rule application from exemplar similarity, none of them may be perfectly correlated with the rule-similarity contrast. The analytic-holistic contrast seems the most diagnostic; it is difficult to imagine true instances of rule application that do not involve acts of selective attention, whereas such acts play no role in most exemplar-similarity models. Differential weighting of attributes also seems like a necessary feature of rule application, but there are important exemplar-similarity models that include this as well (e.g. Nosofsky, 1986; Kruschke, 1992). Similarly, the instantiation of abstract conditions seems to be true of all clear-cut cases of rule following, but again there are exemplar-similarity models that also include it (e.g. Nosofsky et al., 1994). The extensive involvement of working memory is less useful in telling rule application from exemplar similarity, as many cases of rule application may involve only a single condition and consequently place little load on working memory; also, with sufficient practice, there may be little use of working memory in rule application even when the rule involves multiple conditions. Likewise, extensive practice may result in cases of rule application that involve mostly parallel and automatic processing, which would be hard to distinguish from canonical cases of exemplar similarity. Even these less diagnostic distinctions through, are of some use in contrasting many cases of rule application and exemplar similarity.

3. Cognitive studies of rule application vs. exemplar similarity

The cognitive experiments that have dealt with the issues of interest divide into two sets, depending on whether they have employed artificial or natural categories in their research. In what follows, we consider these two sets in turn.

3.1. Studies with artificial categories

3.1.1. Demonstrating and characterizing the basic mechanisms

A useful starting point is an experiment by Allen and Brooks (1991), which is worth describing in detail because it sets the stage for much of what follows. The subjects' task was to categorize imaginary animals into two categories, referred to as 'Builders' and 'Diggers.' Examples of the animals are given in Fig. 1. There were two phases to the experiment: a training phase, during which subjects learned to correctly categorize a set of ten animals, and a test phase, during which subjects were tested on some novel animals as well as on some that they had learned. In the training phase, one group of subjects was taught an additive rule that would distinguish the Builders from the Diggers: e.g. 'If an animal has at least two of the

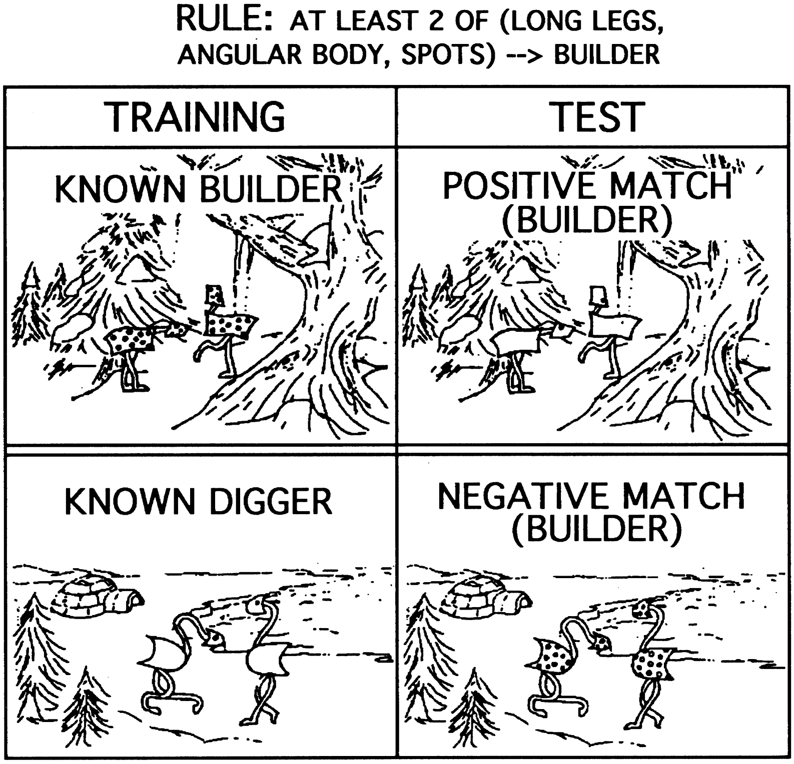

Fig. 1. Examples of materials used by Allen and Brooks (1991). The top right cell illustrates a Positive Match because, not only are these animals Builders according to the rule, but they are also most similar to the studied Builders in the top left cell. The bottom right cell illustrates a Negative Match: though the rule specifies that these animals are Builders, they are most similar to the studied Diggers in the bottom left cell.

following three critical (attribute) values—long legs, angular body, spotted covering—it is a Builder; otherwise, it is a Digger'. A second group of subjects was presented the same animals but was not given the rule. They were told that the first time they saw an animal they would have to guess whether it was a Builder or a Digger, but on subsequent trials they would be able to remember what it was. Thus, the first group was induced to use a rule strategy for categorizing the animals whereas the second group was induced to use a memory procedure. Differences between these 'Rule' and 'Memory' groups speak directly to our questions about differences between rule application and exemplar-similarity procedures.

In addition to the difference between the Rule and Memory groups, the major variation in this experiment concerned the types of items presented during the transfer phase. Two kinds of novel items were of particular interest; it is convenient to illustrate them with respect to the category Builders. One kind of novel item was an instance of Builders according to the rule, and was also extremely similar to an old item that was a known exemplar of Builders (it differed from the known exemplar on only one attribute—see Fig. 1). This kind of item is referred to as a ‘positive match’. The other kind of novel item was also a Builder according to the rule, but it was extremely similar to a known exemplar of Diggers (see Fig. 1); this kind of item is a ‘negative match’. If the Rule subjects do indeed categorize the test items by the rule, their dominant categorization of both positive and negative matches should be the same: Builders. If the Memory subjects categorize a test object by first retrieving the stored exemplar most similar to it and then selecting the category associated with that exemplar, their dominant categorizations of positive and negative matches should differ, with positive matches being labeled Builders and negative matches Diggers. Thus the Rule and Memory groups should differ on their dominant categorization of negative matches[3].

This is just what happened. For negative matches, such as the one illustrated in Fig. 1, the dominant categorization in the Rule group (55%) was Builders, whereas the dominant categorization in the Memory group (86%) was Diggers. These results support the existence of two distinct categorization procedures that can be applied to the same categories corresponding to rule application and exemplar similarity.

Though Builders was the *dominant* categorization in the Rule group, the fact that this rule-based decision occurred only 55% of the time suggests that rule application was not the only procedure at work. Indeed, further analysis by Allen and Brooks (1991) of just the Rule group showed that exemplar similarity was also involved. The logic of their analysis was as follows. If the Rule subjects always applied their rule and never engaged an exemplar-similarity procedure, they should have performed the same on positive and negative matches. But if the Rule subjects sometimes used exemplar similarity as well as rule application, their performance should have been poorer on negative than positive matches; this is because for negative matches, rule application points to one category whereas exemplar similarity points to the other. The latter pattern of results was obtained: the error rate (where an ‘error’ means going against the rule) was about 20% for positive matches, but 45% for negative matches.

Given that the Rule subjects sometimes used both procedures, did they apply the two procedures on the same trial? The reaction-time results for correct responses (i.e. categorizations in accordance with the rule) suggest that the answer is, ‘yes’. Reaction times were longer to correctly-responded-to negative matches than to correctly-responded-to positive matches. This difference fits with the idea that the two procedures were used on the same trial, because extra time would be needed with negative matches to resolve the conflicting categorizations indicated by the two procedures. These results are further compatible with the idea of parallel application of the two procedures.

The picture that emerges from this experiment is in line with much other research on categorization and reasoning (see, e.g. Smith et al., 1992). Rule application and exemplar similarity seem to be distinct procedures that can operate on the identical contents (perhaps in parallel) and sometimes lead to conflicting categorizations. In

[3]Because each positive or negative match was constructed to be highly similar to just one old item, we assume that only that old item is retrieved. Consequently, categorization should be determined by just that retrieved exemplar.

addition, because the outcome of the exemplar-similarity procedure was able to intrude on the processing due to rule application, the former procedure seems to operate more quickly (at least in some circumstances). This latter claim is further supported by the finding that the Rule group took 250 ms longer to respond than did the Memory group.

3.1.2. Triggering conditions

Given the evidence for two procedures, what aspects of a categorization situation trigger rule application versus exemplar similarity? In the study by Allen and Brooks (1991), it seems obvious that the rule instructions induced in subjects a conscious intention to apply the given rule, but it is difficult to go beyond this rather abstract statement.

We can be more analytic about the triggering conditions for the exemplar-similarity procedure. Because this procedure was used even when subjects were trying to employ a rule, presumably the procedure was automatically activated. That is, as earlier suggested in our schematic model of exemplar similarity and discussion of its components (Section 2), the presentation of a test object automatically activates the representations of similar objects. Hence, the triggering conditions for exemplar similarity should include conditions that foster automatic memory retrieval of the relevant object representations. One such condition is that different aspects of an object be integrated because this will lead to the development of a unitary representation, thereby lessening the demands on retrieval processes. Another such condition is that test objects be perceptually distinct, leading to a reduction in interference from other items during retrieval. Follow-up studies by Allen and Brooks (1991) and Regehr and Brooks (1993) support both of these predictions. When the test objects were made either less integrated or less distinctive, there was a decrease in the difference in errors between negative and positive matches—the litmus test for exemplar similarity.

One obvious determinant of automatic memory retrieval has been explored extensively in the memory literature, namely, sheer familiarity with the items (e.g. Shiffrin and Schneider, 1977). When subjects are still learning to categorize test objects, retrieval of a representation of the entire object and its category label will be imperfect (let alone non-automatic), and consequently exemplar retrieval should play little role in categorization, even if no rule is given and subjects are induced to rely on memory mechanisms. These conditions correspond to the early part of the learning phase of many categorization studies, and under these conditions subjects try to generate simple rules to handle their categorization task. These rules are almost always faulty, yet subjects appear to persist with them until automatic memory retrieval starts to take over (Regehr and Brooks, 1993).

This last observation indicates that we may have found a triggering condition for rule application other than sheer instruction. Suppose that people are given a set of objects that assume different values along the same set of separable and salient attributes (standard operating procedure in categorization studies with artificial materials), and are instructed to categorize these objects into a small number of categories. Subjects may naturally selectively attend to the salient attributes, and seek rules that connect combinations of these attribute values to the different categories.

In summary, the following picture emerges for categorization in a task, such as that of Allen and Brooks (1991) in which subjects have been induced to rely on their memories. Early in learning, subjects search for simple rules that can be used to predict the categorizations. This quest for simple rules typically proves futile because of the structure of the materials (e.g. no single attribute is more than 75% predictive of correct categorization) and because the rule that underlies the desired categorizations typically is not obvious (e.g. 2 of 3 critical attribute values). While the quest continues, subjects inadvertently memorize the exemplars and their associated category labels, and eventually the exemplar-similarity procedure is capable of producing correct categorizations, and takes control of performance. Hence, the similarity procedure takes longer to become effective than the rule procedure, though once both procedures are operative it is not unusual for the similarity procedure to operate faster.

The early reliance on simple rules that we see in categorization tasks also occurs in 'free sorting' tasks. In these tasks, subjects are presented a substantial set of objects and asked to sort them into categories (no mention is made of rules). In the experiments of interest (e.g. Medin et al., 1987; Ahn and Medin, 1992), the materials are such that any simple rule (a single attribute value) will not work, in that some of the items will remain unclassified. Subjects use a simple rule anyway, and then sort the remaining items on the basis of their similarity to the instances already sorted by the rule. Again, classification involves two procedures, rules and similarity, with the rule procedure being applied first.

3.1.3. Other dissociations between rule and memory procedures

Consider again the initial finding by Allen and Brooks (1991) that Rule and Memory subjects categorized negative matches differently, with Memory subjects assigning these objects to the same category as their closest neighbors, and Rule subjects assigning them mostly in accordance with the rule. This finding can be viewed as a *dissociation* between the rule-application and exemplar-similarity procedures. Researchers, particularly in neuropsychology, routinely take dissociations as evidence that distinct processes are involved; we draw the same conclusion here. There are several other findings in the categorization literature as well that can be interpreted as dissociations between these two kinds of procedures.

In a series of studies by Smith and Kemler (1984), test objects were squares that varied on two attributes, brightness and side-length. In one experiment, subjects were presented three such objects and instructed to remove the one that did not belong to the same category as the other two. Referring to the three objects as A, B, and C, A and B were alike in that they had the identical value on one of the attributes; in contrast, A and C were the most holistically similar, i.e. the magnitude of their differences summed over both attributes was less than that of A and B. Thus, there were two means for categorizing the objects, on the basis of a common attribute value (e.g. 'smallest side-length'), or on the basis of overall similarity. During the training phase of the experiment, subjects were given feedback that induced them to categorize by one of these two means. There followed a test phase, during which subjects either had to make their categorizations very rapidly or not. The major result was that at fast speeds the two categorization strategies led to equally accurate performance, whereas at slower speeds categorization based on a common attribute was more accurate. Thus, speed of responding dissociated the two categorization strategies.

The connection of these results to our concerns is straightforward. Categorization based on a common attribute requires selectivity attending to that attribute and determining if the attribute's value instantiates some abstract condition. These are two of the key components of rule application, suggesting that categorization based on a common attribute amounted to applying the rule, 'If two objects have the smallest line length, then they're in the same category'. In contrast, categorization based on overall similarity involved a similarity computation like that used in the first stage of the exemplar-similarity procedure. The dissociation between the strategies further suggests that the successful use of rule application required more time than that needed for exemplar similarity, perhaps because the former procedure involves more time-consuming processes.

Another experiment by Smith and Kemler (1984) involved the same objects and tasks as those described above, except that the rule- and similarity-based categorization strategies were induced by instructions (roughly, 'Respond on the basis of first impressions' versus 'Carefully decide') rather than by feedback. Similarity-based categorizations were faster than rule-based ones. Again, speed of responding dissociated rule and similarity procedures. Still another study in this series varied whether subjects using either rule- or similarity-based strategies had to perform a concurrent task or not. Performing a secondary task interfered with categorization more when subjects were categorizing by rule. So a dual-task requirement also dissociated rule and similarity procedures. Using verbal materials rather than pictorial ones, Smith and Shapiro (1989) have also found that it is easier to perform a secondary task when one is categorizing by similarity than by rule.

3.1.4. Implications of findings

These dissociations fit well with our previous discussion of the components of rule application and exemplar similarity. The findings that a rule-based strategy typically operates slower than a similarity-based one and is more disrupted by a secondary task suggest that the rule strategy demands more of some time- or effort-limited cognitive component than does the similarity procedure. Two such components implicated by the schematic models that we sketched earlier are selective attention and working memory.

These ideas about component processes are summarized in Fig. 2, which fleshes out our earlier schematic models of rule application and exemplar similarity. Note that rule application involves: selective attention (Boxes 2 and 4), making perceptual tests that correspond to conditions of the rule (Box 5), and extensive working-memory operations (Boxes 3, 6 and 7). In contrast, exemplar similarity involves: retrieval from long-term memory (Boxes 2 and 3), and a subsequent similarity-comparison process (Boxes 7 and 8). These models are compatible with the cognitive findings that we have reviewed thus far, and as we will see, with the neurological findings that we present below.

3.2. Studies with natural categories

3.2.1. Studies demonstrating the two categorization mechanisms

In the experiments reviewed thus far, the relevant attribute-values were readily

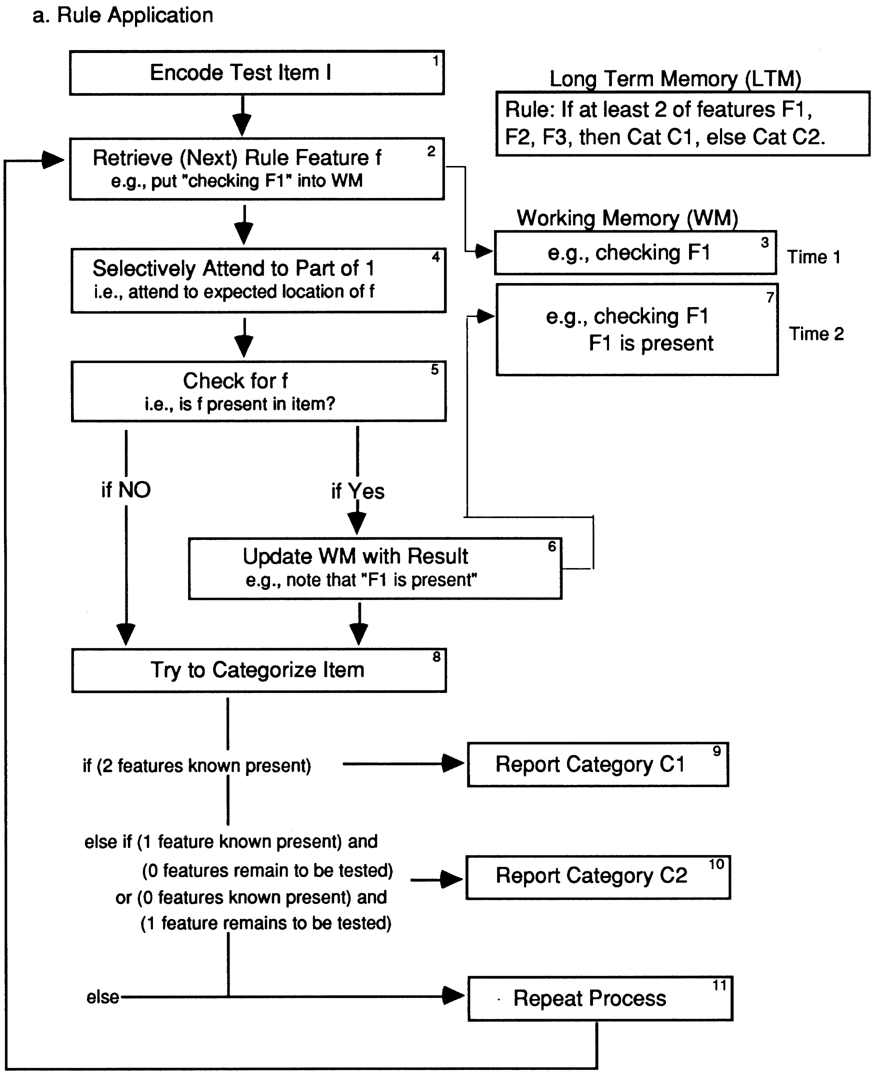

Fig. 2. (a) A model of the rule-application process as applied to Allen and Brooks (1991) categorization task; (b) a model of the exemplar-similarity process as applied to the same task. In (a) the term 'feature' designates an attribute value.

separated from one another and were quite salient. This is typical of studies with artificial categories. In many studies of natural categories, though, the test object is either a picture of a familiar object or a frequent word, and the relevant attribute-values (either in the picture or the word's meaning) seem to be less salient. Hence, most experiments with natural categories are not strictly comparable to most studies with artificial materials. However, a number of researchers have recently emphasized a new paradigm for studying natural categories that more closely parallels experiments with artificial materials. In this paradigm, subjects are presented with a partial textual description of an object (e.g. 'a small animal that burrows in the earth...') and subjects have to decide whether or not that object is a member of some target category. Because the relevant attribute-values are denoted by distinct words in the description, they are readily separable and salient, just as is the case with artificial materials (though the attributes are semantic rather than perceptual). All of the natural-category studies that we consider are of this type.

A developmental study by Keil and Batterman (1984) speaks to the existence of different categorization strategies. The experiment involved children aged 5–10 years. On each trial, subjects were presented descriptions of two items and had to decide which one belonged to a target category. The categories were natural ones,

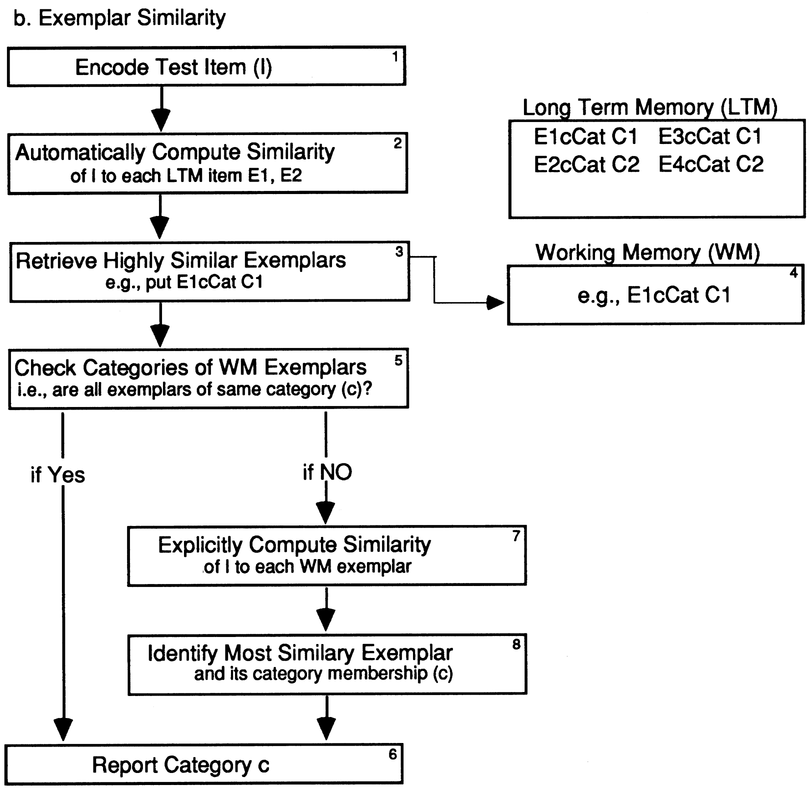

Fig. 2b.

and each contained features (attribute values) that were *necessary* (i.e. true of all members) as well as features that were only *characteristic* (i.e. true of typical members) (Smith et al., 1974). One of the descriptions presented on a trial contained the necessary features of the target category but not the characteristic ones, whereas the other description contained the characteristic features but not the necessary ones. For example, when the category was Robber the two descriptions were:

1. A very friendly and cheerful woman who gave you a hug but then disconnected your toilet bowl and took it away without permission and no intention to return it (necessary but not characteristic); and
2. A smelly mean old man with a gun in his pocket who comes to your house and takes your TV set because your parents didn't want it any more and told him he could have it (characteristic but not necessary).

The major findings were that younger children were more likely to select descriptions with characteristic features, whereas the older subjects favored descriptions with necessary features.

To connect these results to the concerns of the present paper, we note that a description with characteristic features seems more similar to exemplars of its category than does a necessary-feature description. Thus, selection of a characteristic-feature description might well be based on the exemplar-similarity procedure (or on a similarity-to-prototype procedure). In addition, let us assume that selection of a necessary-feature description is often based on noticing the diagnostic status of the necessary feature, selectively attending to it, and then using it as a rule (e.g. 'If someone took someone else's goods without permission..., they're a robber'). Given these assumptions, the results from Keil and Batterman (1984) imply that younger subjects relied more on exemplar similarity than rule application, whereas older subjects did the reverse. This is a dissociation between exemplar-similarity and rule-application procedures, and hence these results converge with those obtained with artificial categories.

However, because these results involve a contrast between two groups of subjects, they do not demonstrate that the same subject applied the two procedures to the same item, as in some of the experiments with artificial materials. A recent report by Hampton (1995) comes closer to offering this demonstration. Hampton's materials were very similar to those of Keil and Batterman (1984), but he used adult subjects. On each trial, subjects were presented a category and a description of an object, and they had to decide whether or not the object was a member of the category. Several kinds of descriptions were used, and four of them are of particular interest. Relative to the category with which it was paired, a description could contain:

1. both the necessary and characteristic features;
2. the necessary but not the characteristic features (like the necessary-feature descriptions of Keil and Batterman, 1984);
3. the characteristic but not the necessary features (like Keil and Batterman's characteristic-feature descriptions); and
4. neither the necessary nor the characteristic features.

Consider first the results for conditions (3) and (4). Because no necessary feature is present in either case, there is no possibility of rule application. If subjects rely on a similarity mechanism (either exemplar similarity or prototype similarity), there should be more positive categorizations (‘yes, it’s a member’) in (3) than (4). This is what Hampton (1995) found. Now consider conditions (2) and (4). Because a necessary feature is present in (2) but not (4), rule application will support a positive categorization only in (2), and hence there should be more positive categorizations in (2) than (4). Again, this is what Hampton found. Neither of these results is surprising. Of greater interest is the contrast between conditions (1) and (2). Because a necessary feature is present in both conditions, rule application will support a positive categorization in both conditions. If only rule application is used, the two conditions should not differ in the frequency of positive categorizations; but if exemplar-similarity is also activated (perhaps automatically), there should be more positive categorizations in (1) than (2). Indeed, this is what Hampton (1995) found, confirming that both categorization procedures could have been active. Note that this pattern of results is very similar to those of Allen and Brooks (1991) who used artificial categories[4].

The convergence of results between studies with natural and artificial materials, however, rests on the assumptions that we have made about when subjects use similarity and rule procedures in natural categorization. It would be useful to ‘ground’ these assumptions; i.e. to provide evidence that subjects are computing similarity when the described object contains features characteristic of the target category, and are using a rule when focusing on necessary features. A procedure introduced by Rips (1989) provides some empirical confirmation.

Rips (1989) was primarily interested in demonstrating that categorization is not always based on similarity. In his paradigm, on each trial a subject was presented a description of a test object that mentioned only a value of a single attribute (say, the object’s diameter). Then the subject had to decide to which of two target categories the object belonged (it had been previously established that the object was between the subject’s extreme values for the two categories). For example, one description might be, ‘A circular object with a 3-in diameter’, with the associated categories being Quarter and Pizza. The description was constructed to be smaller than the smallest pizza in that subject’s experience, but larger than the largest quarter. Now, quarters are restricted in size but pizzas are not, and if subjects brought this piece of knowledge to bear they would likely decide that the object is a pizza. All of the items had this structure: one category–the ‘variable’ category—always allowed more variability on the relevant attribute than did the other category—the ‘fixed’ category. This variation is similar to varying whether a necessary feature is present or not. What is novel about the paradigm of Rips (1989) is that, in addition to asking subjects for categorization decisions, he also had them rate the similarity of each

[4]It is worth noting that Hampton (1995) interpreted his results differently. He took the finding of more positive categorizations in condition (1) than (2) to mean that the so called ‘necessary features’ present in conditions (1) and (2) were not really necessary, and that category membership is always a matter of degrees. However, Osherson and Smith (1997) present arguments against Hampton’s interpretations.

described object to the target categories. This provides a clear indication of which category would be favored by a similarity mechanism.

Rips (1989) found that subjects were more likely to categorize the test objects as members of the variable category (Pizza in our example), but rated the objects as more similar to the fixed categories (Quarter). So we can conclude that categorization was not based on a similarity computation. What was it based on? Presumably, it was based on a rule that focuses on the constraint of the fixed category, e.g. ‘If an object is more than 1-in in diameter, it cannot be a quarter.’ Given this assumption, Rips’ results show a dissociation between similarity- and rule-based judgments. What they do not show, though, is a dissociation between two categorization procedures, because the similarity procedure was operative only in similarity judgments. Also, there is nothing in Rips’ results that grounds our assumption about rule use by showing that a focus on a necessary feature involves the application of an explicit rule.

What is needed is a variant of the Rips paradigm—collecting both similarity and categorization responses—that in some cases fosters similarity-based categorizations and in others fosters rule-based categorizations. Smith and Sloman (1994) report just such an experiment. In some conditions of their experiment, they essentially replicated Rips’ study—e.g. Is a circular object 3 in in diameter more likely to be a Pizza or a Quarter? Is it more similar to a Pizza or a Quarter? The results in these conditions replicated those of Rips in that subjects judged the objects as more likely to be members of the variable categories but more similar to the fixed categories. In other conditions, the descriptions were enriched so that they contained features characteristic of the fixed categories but not of the variable categories, e.g., ‘A circular object with a 3-in diameter that is silver colored’. In these conditions, subjects judged the described objects not only as more similar to the fixed categories, but also as more likely to be members of the fixed categories. That is, categorization judgments tracked similarity judgments. This provides strong evidence that the categorization judgments were indeed based on a similarity procedure.

Finally, consider the evidence for our assumption linking a focus on necessary features to the explicit use of rules. Smith and Sloman (1994) had subjects ‘think aloud’ while making their decisions, and subsequently analyzed these verbal protocols. In those condition that used the sparser descriptions (e.g. ‘A circular object with a 3-in diameter’), on some trials subjects would justify their categorization decision by explicitly stating the relevant rule, e.g. ‘Quarter can’t be larger than 1-in’. In such cases, subjects chose the variable category (e.g. Pizza) all of the time. The explicit statement of a rule is a standard criterion for rule use (Smith et al., 1992), and the all-or-none responding fits with the all-or-none nature of rule use.

3.2.2. Implications of the findings

These results converge nicely with the findings obtained with artificial materials, particularly the work of Brooks and his colleagues. Even when a rule is present that can be used for categorization, subjects will also invoke a similarity procedure if the test object contains information that is characteristic of one of the target categories.

The results with natural categories are also roughly compatible with the models presented in Fig. 2. Rule application with verbal materials again involves selectively attending to relevant attributes, though now the attributes are semantic (one's knowledge about the categories of interest). In our Pizza-Quarter example, presumably subjects inspect their semantic representations for pizzas and quarters, selectively attending to information about size. Rule application in these semantic cases may also sometimes involve a kind of instantiation process; e.g. one's knowledge about apples includes their distinctive coloring, but one must decide if that distinctive coloring is adequately captured by the word 'red' in the description 'A circular object with a red color.' What is less likely is that working memory plays much of a role in the natural-category experiments. With natural, semantic categories, one does not have to rehearse the rule (necessary feature), nor keep track of how many attributes have been checked thus far, but there may still be a need for keeping some information in an activated state.

With regard to the exemplar-similarity procedure, again the results with natural categories are compatible with the schematic model in Fig. 2. In particular, when a description contains a sufficient number of characteristic features, the category of the described object seems to be automatically activated.

4. Neuropsychological and neuroimaging studies of rule application vs. exemplar similarity

Another way to determine whether rule application and exemplar similarity are qualitatively different strategies is to ascertain whether they are implemented by different neural structures in the brain. In what follows, first we consider some indirect evidence from neuropsychology (i.e. the study of selective deficits due to brain damage), and then present some direct evidence from a neuroimaging experiment from our laboratory.

4.1. Neuropsychological evidence for qualitatively different procedures

The central assumption underlying the exemplar-similarity procedure is that categorization rests on previously stored examples. Consider patients who have damage in regions of the brain known to be involved in storing new items: such patients should have difficulty storing exemplars of a new category, and consequently have difficulty using the exemplar-similarity procedure in future classifications involving this category. This prediction has been assessed by Kolodny (1994). He tested patients with damage in their medial-temporal lobes, which contains the hippocampal system that is known to be critically involved in the consolidation of new memories (e.g. Squire, 1992). Kolodny compared such patients to normal controls on two categorization tasks. One task required subjects to learn to sort novel paintings into two categories that corresponded to two different artists, whereas the other task required subjects to learn to sort dot patterns into two categories that corresponded to two different prototypes. Independent behavioral evidence indicated that only the paintings task typically recruits an exemplar-similarity procedure in normal subjects (the dot-patterns task seems to trigger a reliance on abstract prototypes). The patients learned the dot-pattern categories as readily as the normal controls, but performed far worse than normal on the painting categories. This leads to the inference that damage to the medial-temporal lobe selectively impairs categorization based on exemplar similarity.

With regard to the neural basis of rule following, among the neuroanatomical areas likely to be involved are the frontal lobes. Clinical observations have long suggested that damage to this region is associated with difficulties in thinking analytically and applying abstract rules (e.g. Luria, 1969). Also, there are many experiments which demonstrate that patients with frontal-lobe damage perform substantially less well than normal controls on tasks that require the use of explicit rules. The task of choice is typically the Wisconsin Card Sort Task. On each trial, a card is presented that contains colored geometric forms; from card to card there is a variation in the number, shape, size and background shading of these forms. The subject must learn which one of the four attributes to use as a basis for sorting the cards into four piles. Once subjects have sorted a certain number of cards correctly, the experimenter switches the relevant attribute and subjects now have to discover the new critical attribute. Frontal-lobe patients are strikingly impaired on this rule-based task, not only compared to normal subjects but compared to patients with brain damage outside of the frontal lobes. The frontal-lobe patients may learn the initially relevant attribute as well as other subjects, but they have severe problems shifting to a new rule when the experimenter switches relevant attributes (e.g. Milner, 1964).

The preceding studies are suggestive, but they have definite weaknesses when it comes to providing strong evidence about qualitative differences between rule- and similarity-based categorization. For one thing, these studies deal more with the acquisition of novel categories than they do with categorization using already-learned categories, and it is the latter topic that has been the main concern in this paper. Another matter is that no published experiment has tested two patients of interest—say, a frontal-lobe and a medial-temporal-lobe patient—on two tasks of interest—one that recruits primarily rule application and one that relies mainly on exemplar similarity—showing that one patient is impaired on one task but not the other, whereas the other patient shows the reverse pattern. That is, no neuropsychological double dissociations between rule application and exemplar similarity have been demonstrated.

However, there is a recently obtained neuropsychological double dissociation in the area of lexical processing, and it is relevant here (Ullman et al., 1997). The experiment of interest is based on the prior research of Pinker and colleagues on forming the past tenses of regular and irregular verbs (e.g. Pinker, 1991). This work indicated that generating the past tense of a regular verb (e.g. 'jump'-'jumped') is done by application of a rule (roughly, 'Add -ed to the present-tense'), whereas generating the past tense of an irregular verb (e.g. 'sing'-'sang') is accomplished by retrieving relevant information from memory, including information about similar exemplars (e.g. 'ring'-'rang'). These two procedures have an obvious similarity to rule application and exemplar similarity. Ullman et al. (1997) had patients convert regular and irregular present-tense verbs to their past tenses. They found that patients with Parkinson's disease, who are known to have problems in applying systematic procedures, made more errors on regular than irregular verbs; in contrast, Alzheimer's patients, who have well-documented memory problems, had more difficulty with irregular than regular verbs. This establishes a double dissociation between rule application and exemplar retrieval in a task related to categorization. Furthermore, there is some neuroimaging evidence for this double dissociation, as different brain areas are activated when normal subjects convert regular and irregular present-tense verbs to their past tenses (Jaeger et al., 1996).

4.2. Neuroimaging evidence for qualitatively different procedures

4.2.1. Rationale for the experiment

In an effort to provide more direct evidence for different neural correlates of rule application and exemplar similarity, we conducted an experiment in which normal subjects performed categorization tasks while their brains were scanned using positron emission tomography, or PET (a technique for measuring changes in regional cerebral blood flow as an index of changes in regional neural activity). One group of subjects performed a rule-based categorization task while another performed an exemplar-based task. To the extent that different regions of the brain are activated in the two tasks, we have evidence that qualitatively different cognitive procedures are involved. To the extent that the known functionality of the activated areas in a task corresponds to processes that are thought to be involved in the strategy underlying that task, the PET evidence is strong indeed[5].

The categorization tasks we used were variants of those used by Allen and Brooks (1991). Recall that in that experiment a Rule group and a Memory group first learned to categorize imaginary animals, and then were tested on new animals in a transfer phase. The results from the transfer phase showed that the Memory group relied on an exemplar-similarity procedure, whereas the Rule group relied primarily on rule application. These tasks are well-suited for a PET study for two reasons. Firstly, in a PET study one needs to ensure that differences in brain activity cannot be attributed to differences in the complexity or familiarity of stimuli, and the only difference between Allen and Brooks (1991) two tasks is in the instructions. In all other respects—including amount of training provided prior to testing, the nature of the test items, the number of response alternatives, etc.—the two tasks are identical.

[5]Basically, PET works as follows. It is known that regional neural activity in the brain causes regional increases in blood flow. A radioactive substance (^{15}O) is injected into a subject's bloodstream, and flows to the parts of the brain that are neurally activated during the task. As the substance decays in the brain, it emits positrons. Each positron moves only a few mm before it collides with an electron. The annihilation process produces two photons that travel outward from the point of collision in opposite directions. A PET scanner contains rings of photon detectors surrounding the subject's head. When two photons are detected on opposite sides of the detector at nearly the same moment, they are assumed to have come from the same annihilation process. Using tomographic techniques, it is then possible to construct images of where in the brain the annihilations occurred and, by inference, where in the brain there was neural activation.

Secondly, there seems to be little doubt that the Rule and Memory tasks do indeed involve rule application and exemplar similarity, respectively, whereas this difference in procedure is less clear in any of the paradigms that used natural categories.

There is, however, a problem with this choice of tasks. As noted earlier, subjects in the Rule group seemed to make some automatic use of exemplar similarity. This conclusion is based on the fact that, during the transfer phase, the Rule group responded more slowly and less accurately to negative matches—new test animals that were highly similar to an old animal in the wrong category—as compared with positive matches—new test animals that were highly similar to an old animal in the correct category. We, nonetheless, chose to use the Rule task for the following reasons. Firstly, some automatic use of exemplar similarity is unavoidable unless one is willing either to: (a) present different items to the Rule and Memory groups in the transfer phase, such that the Rule subjects see test objects that are highly dissimilar to the training ones and therefore are unable to use exemplar-based processes; or (b) provide different amounts of training to the two groups, such that the Rule subjects would have less practice than the Memory subjects, and therefore have less time to learn the study items in the first place. Either of these changes would seriously compromise our ability to interpret the PET results. Secondly, there is reason to believe that exemplar similarity plays considerably less of a role in the Rule condition than does rule application. In Allen and Brooks (1991), the difference in accuracy between positive and negative matches—the litmus test for the use of exemplar similarity—was approximately 25% in the Rule condition as opposed to approximately 60% in the Memory condition. Hence, even if subjects do engage exemplar-based processes in the Rule condition, the impact of these processes should be relatively small and contribute little to increases in brain activation, especially in the context of the demanding and time-consuming rule-application processes. Finally, even if both memory- and rule-based processes turn out to be responsible for regional blood-flow changes in the Rule condition, those activations that are due to memory processes should be detectable by spotting brain regions active in both the Rule and Memory conditions.

4.2.2. Procedure

We altered the Allen and Brooks (1991) tasks in a number of respects (Smith et al., submitted). Some changes were made to maximize reliance on memory in the Memory condition and time spent on rule application in the Rule condition, thereby improving the chances that these procedures would be captured by PET. To maximize reliance on memory, the imaginary animals were created with more perceptually distinctive features, given that we have already noted that perceptual distinctiveness increases memory use (Regehr and Brooks, 1993). Also, the study period was extended to ensure that the animals were well-learned. To maximize time spent on rule application, we increased the complexity of the rule (it now required matches on at least 3 of 5 attribute-values). Another change was that subjects were given a 2.5-s time window in which to make a response rather than having trials scheduled ad lib. We did this to ensure that the number of categorization decisions was equated across conditions, so that it could not be the cause of differences in PET results. Finally, the experiment was conducted over 2 days. On the first day, subjects performed four blocks of study trials, with two repetitions of the ten study stimuli in each block, followed by one block of transfer trials. In the second session, subjects completed one more block of study trials, and were then scanned while doing three blocks of transfer trials.

In addition to engaging in either the Rule or Memory condition, subjects performed three blocks of a control condition while being scanned. The purpose of this condition was to capture some of the processes that were presumably operative in the Rule and Memory conditions but were not part of categorization per se, and hence not of direct interest. Such irrelevant processes include basic perceptual and response processes. Following standard PET methodology (Posner et al., 1988), the regions activated in this Control task were subtracted from those activated in the Rule and Memory tasks, in order to isolate brain regions associated only with the categorization processes of interest. The Control task was the same for both Memory and Rule subjects, and simply required subjects to push either of two buttons at random whenever a test object appeared. The pictures and response buttons were the same as those used in the categorization conditions.

4.2.3. Results

Consider first the behavioral findings. We focus here on just the results collected during the PET scans, though the results from the first session are quite similar. By and large, the results replicate those obtained by Allen and Brooks (1991). The Rule group took considerably longer than the Memory group to make their responses—an average of 760 ms longer. Furthermore, subjects in the two groups differed in their dominant categorization of negative matches—those items that followed the rule of one category but were most similar to a studied item in the other category. Rule subjects categorized 71% of these items according to the rule, whereas Memory subjects categorized a full 76% in terms of similarity to study items. These results replicate the basic dissociation between the Rule and Memory groups, and suggest that subjects in the two conditions used different processes to categorize the items.

Like Allen and Brooks (1991), we found evidence that the Rule group was also influenced by exemplar-based processes. Specifically, Rule subjects were more accurate on positive (85%) than negative matches (71%). Unlike Allen and Brooks (1991), however, we found no effects of positive vs. negative match on response times. This may have been a result of the time-window procedure employed here; i.e. the procedure may have encouraged subjects to respond at a consistent speed for fear of failing to respond before the deadline. Taken together, these results show that, though there was some use of exemplar similarity in our Rule condition, it was relatively minor; the accuracy difference between positive and negative matches (14%) was small compared to that obtained in the Memory condition (53%), and was even less than that obtained in the Rule condition of Allen and Brooks (25%). These comparisons suggest that Rule subjects mainly did rule processing, which means the PET results should be relatively pure.

The PET results for each condition were computed by subtracting the Control activation data from the Rule or Memory activation data for each subject, and then

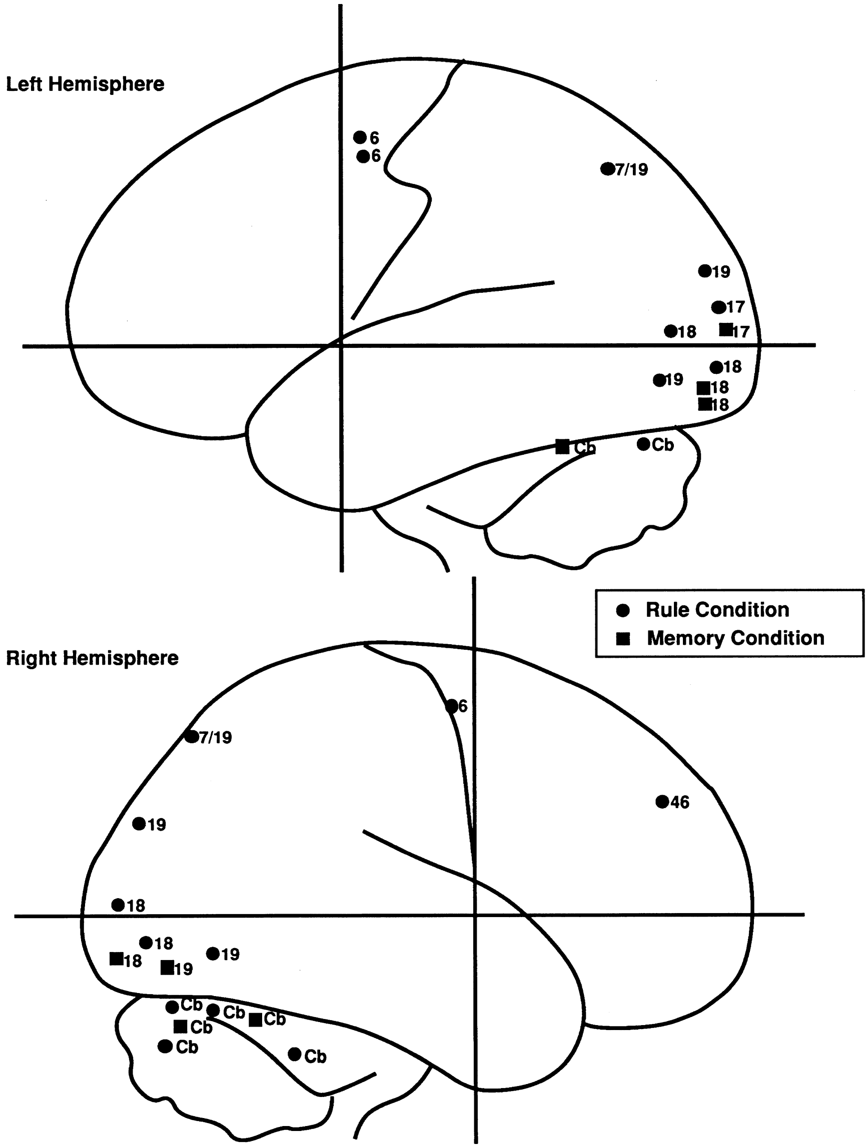

Fig. 3. Marked areas on right and left lateral surfaces of the cortex denote regions of significant activation in Rule- and Memory-condition subtraction images. Numbers correspond to Brodmann's areas; Cb, cerebellum. Active regions not visible from this perspective are: left Brodmann Area 31 (it is on the medial surface of the parietal lobe and is active in the in the Memory condition), and the right thalamus (it is a subcortical structure active in the Rule condition).

averaging the subtraction data across all subjects within each condition. The resulting images revealed numerous areas of significant activation; these are presented schematically in Fig. 3. The main finding is a dissociation: Of the 23 areas activated across both conditions, 14 were activated solely in the Rule condition. Of the remaining nine areas, seven were active in both conditions[6]; only two were active in the Memory condition alone. All of these areas are labeled on the schematic using numbers designated by the physiologist Brodmann to refer to regions of the cortex differing in cytoarchitecture (as well as the designation ‘Cb’ to refer to areas in the cerebellum).

Areas of activation distinct to the Rule condition are as follows: two areas at the top of the parietal lobes (one in each hemisphere; Brodmann Area 7 in both cases); areas in the frontal cortex, namely, one area in right dorsolateral prefrontal cortex (Brodmann Area 46) and three areas in supplementary motor cortex (two in the left hemisphere and one in the right hemisphere of Brodmann Area 6); four areas in the visual cortex (two in left Brodmann Area 19, one in right Brodmann Area 19, and one in right Brodmann Area 18); three areas in the cerebellum (two in the right and one in the left); and one area in the right thalamus (not shown in the figure because it is a medial structure and only lateral views of the brain are illustrated).

Areas of activation common to both Rule and Memory conditions are: five areas in the visual cortex (one in left Brodmann Area 17, one in each hemisphere of Brodmann Area 18, one in left Brodmann Area 18/31[7], and one in right Brodmann Area 19), and two areas in the right cerebellum. The remaining two areas (one in left Brodmann’s Area 18 and one in the left cerebellum) were active only in the Memory condition. At a minimum, the entire pattern of results suggests that there are mechanisms involved in rule-based categorization beyond those involved in exemplar-based categorization, which supports the proposal that rule-based categorization is qualitatively different from exemplar-based categorization. As we are about to see, these additional rule mechanisms are exactly the ones implicated in our previous discussion of the cognitive literature.

Further evidence for qualitative differences between rule and exemplar procedures comes from functional interpretation of the activated areas. Firstly, consider the areas distinctive to the Rule condition. The largest activation is that in the left-hemisphere parietal lobe (Brodmann Area 7). In single-cell studies with non-human primates, neurons in this area of parietal cortex have been found to fire when the organism must selectively attend to certain locations (Bushnell et al., 1981). And studies with human neurological patients show that damage to this region of parietal cortex is associated with impairments in spatial, selective attention. As we have

[6]Each area of activation contains a peak—a point of greatest change in activation within the area—that can be specified using an x, y, z coordinate system. An area of activation found in the Rule condition is considered ‘common’ to an area found in the Memory condition if peaks of the two areas differ by less than 10 mm—the spatial resolution of PET—on each coordinate.

[7]The peak is in Brodmann Area 18 in the Rule condition and in Brodmann Area 31 in the Memory condition. The latter is not shown in figure because Area 31 is a medial structure.

[8]Actually, the evidence about parietal function pertains to selective attention for spatial position, whereas our earlier discussion of rule application emphasized selective attention for different attributes. There is no real conflict here, though. The relevant attributes of the items in our PET study and in other studies of visual categorization consistently occur in different spatial positions (e.g. ears at the extreme left, tail at the extreme right); hence, attending to a particular attribute entails attending to a particular location.

noted repeatedly, selective attention is part of rule application but not exemplar similarity, and hence the activation at issue fits well with our proposals[8].

Another region activated in the Rule condition was in the prefrontal cortex (Brodmann Area 46). As noted in the previous section, damage in the prefrontal cortex in neurological patients is routinely associated with difficulties in applying rules. Thus, the PET data converge with the neuropsychological results in indicating the importance of the prefrontal region in rule application. But what psychological function does this region serve? One possibility, based on other neuroimaging experiments (e.g. Cohen et al., 1994; Smith et al., 1996), is that the prefrontal cortex is activated whenever operations are performed on material currently maintained in working memory. As argued earlier, such working-memory operations should be more frequent in the Rule than the Memory condition[9].

A third area activated in the Rule condition was in supplementary motor cortex (Brodmann Area 6). In general, this area has been shown to play a role in the high-level preparation and planning of movement (Fuster, 1995), including speech (Petersen et al., 1988). Of specific interest here is that activation of supplementary motor cortex has been reported in studies of verbal working memory, and is believed to mediate implicit speech, or rehearsal, of stored verbal information in these contexts (Awh et al., 1995). This region, then, may have been involved in rehearsing both the rule (e.g. ‘If an animal has a spotted body...’) and the intermediate results of rule application (e.g. ‘has spotted body’) during categorization in the Rule condition.

The other activated regions in the Rule condition include four visual areas that were not significant in the Memory condition, three cerebellar regions, and one area in the thalamus. The former regions may be involved in processes that generate a percept with a salient part structure, which was needed in the Rule condition but not necessarily in the Memory condition. Alternatively, the visual activations may reflect the analytic perceptual tests made in the Rule condition, e.g. ‘Are the legs long?’. Either of these preceding suggestions is speculative (particularly since the areas at issue are somewhat dorsal and posterior to the areas usually associated with the analysis of object parts). We have no firm interpretation of the thalamic or cerebellar activations either, but we note that the latter activations are extremely common in PET studies of working memory (see, e.g. Smith et al., 1996).

Next, consider the seven areas of activation common to the Rule and Memory conditions. Two areas active in both the Rule and Memory conditions were in the right cerebellum; again, we have no firm interpretation of these activations. The remaining five areas were all in the visual cortex. One interpretation of these results is that the region in the visual cortex may be involved in automatic aspects of

[9]There are other views of the function of the prefrontal cortex. One is that it is involved whenever the person has to shift attention and processing from one subtask to another (Rubenstein et al., 1994; D’Esposito et al., 1995). Perhaps the need for Rule subjects to shift from one attribute to another is akin to a task shift. Another proposed function of the prefrontal cortex is that it is responsible for inhibition (e.g. Diamond, 1988; Shimamura, 1994). Inhibition may well have played a role in our Rule condition. Whenever a negative match was presented, its close neighbor from the wrong category may have been activated, and consequently subjects would have had to inhibit responding in accordance with this retrieved exemplar.

exemplar retrieval, since this process is occasionally operative in the Rule condition, and of course is prevalent in the Memory condition. That is, the processing revealed may reflect the storage and retrieval of visual memories, namely, animal stimuli. This interpretation is bolstered by recent PET findings about the neural bases of visual imagery (Kosslyn et al., 1993): retrieving a visual image is a paradigmatic case of retrieving a visual memory, and visual-imagery tasks show activation in visual regions close to those activated here (though the imagery regions tend to be more medial).

A second possibility is that, given that these areas of activation were only *roughly* in the same locations for the Rule and Memory conditions, they might be the result of different kinds of processing in each condition. In the Memory condition, the occipital activation might have mediated retrieval of visual memories (as described in the previous paragraph) and perhaps subsequent computations performed on the retrieved representation (e.g. 'Is it sufficiently similar to the test object?'). However, in the Rule condition, the activation may be due to perceptual testing of rule-relevant stimulus features. A third possibility is that more perceptual processing was required in both the Memory and Rule tasks than in the Control condition, and that the occipital regions in question mediate some of this additional processing.

Finally, there are two areas active in the Memory condition only. Because these areas are in the cerebellum and the visual cortex—regions in which common Rule and Memory areas were also found—it is unlikely that the two areas reflect novel cognitive processes. More likely, they (particularly the area in visual cortex) further mediate uniquely Memory-oriented processes such as the retrieval of stored exemplars.

In summary, the most important PET results are:

1. There is a striking dissociation between the neural regions activated in the Rule and Memory conditions (of 23 significant areas, 14 are distinct to the Rule condition); and
2. The known functionality of the activated regions supports the hypotheses that only rule application involves selective attention and working memory, while visual-perceptual and/or visual-memory processes may be common to both procedures of interest.

The results are consistent with the view that rule-application and exemplar-similarity strategies are indeed qualitatively distinct.

5. Summary and other issues

5.1. Summary

Studies with artificial categories provide evidence that rule application and exemplar similarity are qualitatively different categorization procedures (Allen and Brooks, 1991; Regehr and Brooks, 1993). The most straightforward piece of evidence is that subjects instructed to use a rule differ in predictable ways in their dominant categorization of certain items from subjects instructed to rely on their memories. Other experiments with artificial categories provide further dissociations between conditions that foster rule use and conditions that foster reliance on memory; e.g. rule use is more affected by relative emphasis on speed (Smith and Kemler, 1984). The picture that emerges from all these studies is that, compared to exemplar similarity, rule application is a relatively slow and analytic process, presumably because it involves extensive use of selective attention and working memory.

The limitation of the above studies is that they are quite artificial, presenting highly analyzable materials while explicitly encouraging subjects to rely either on rules or memory. However, this limitation is not true of the experiments with natural categories. These studies showed, for example, that under some conditions categorization was likely based on similarity to stored items, whereas in other conditions categorization could not be based on similarity to stored exemplars and instead presumably depended on rule use (Smith and Sloman, 1994). These studies were limited by their lack of independent evidence for rule use. But the fact that the natural-category results converged so well with the data obtained with artificial categories bolsters the case for two distinct categorization procedures.

This pattern of converging results was strengthened by findings from neuropsychology and neuroimaging. The neuropsychological data suggested that selective brain damage can lead to the selective loss of either rule application of exemplar retrieval (Ullman et al., 1997). The neuroimaging experiment provided further evidence for qualitatively different categorization procedures. Not only were many of the areas activated during rule use distinct from those activated during memory-based categorization, but some of the Rule areas were independently known to be involved in selective attention and working memory, processes that characterize rule application but not exemplar similarity.

The overall package of results seems most parsimoniously explained by positing two distinct procedures that can operate independently of one another, perhaps in parallel. While attempts to handle this evidence by a unitary theory—e.g. it's all due to exemplar similarity—might have some success in dealing with the standard cognitive experiments, it is difficult to envision how a unitary approach can encompass the neurological data as well.

5.2. Other issues

In addition to the distinction between rule application and exemplar similarity, there are numerous other suggested distinctions between categorization processes and representations. A few of these deserve brief mention.

We noted at the outset that in addition to the two categorization strategies on which we focused, there are at least two others that have a good deal of currency. These are prototype similarity and theory application. With regard to the former, to the extent that one construes a prototype to be the single best example of a category (in the tradition of Rosch, 1973), it may be difficult to distinguish a prototype-similarity procedure from an exemplar-similarity one. Still, the neuropsychological results of Kolodny (1994) suggest that damage to the medial-temporal lobe impairs performance on tasks based on exemplar similarity but not performance on tasks based on prototype similarity. This dissociation clearly supports the distinction between exemplar and prototype similarity, and Squire and Knowlton (1995) have produced a comparable dissociation. With regard to theory application, to the extent one construes it as zeroing-in on just the few core features of a concept and using only them for categorization, it may turn out to be indistinguishable from rule application. But if theory application is instead seen either as providing the best possible explanation of all available features of a test object (as in Rips, 1989), or as a kind of ‘proof’ of the test object’s feature from the category’s more abstract attributes (as in Ahn et al., 1992), then it too may turn out to be distinct from other categorization strategies.

While we have focused on the difference between specific categorization procedures, others have considered more abstract processing differences. One such difference is that between analytic and holistic processing. We noted earlier that rule application and exemplar similarity differ with respect to this processing attribute because only the former involves selective attention, but there may be more than selectivity to the analytic-holistic distinction in categorization (e.g. only in analytic processing may an object be decomposed into its parts; see Farah, 1990). Another abstract distinction is that between explicit and implicit categorizations. In the former case, categorization is based on information that is explicitly represented, whereas in the latter case it is based on nonconscious implicit knowledge. Neuropsychological studies have already shown a dissociation between those two kinds of categorization (e.g. Knowlton et al., 1992; Squire and Knowlton, 1995). This explicit-implicit distinction may be relevant in some contrasts of rule- and similarity-based categorization procedures, as a rule is generally assumed to be explicitly represented whereas a retrieved exemplar need not be. However, in some of the tasks we have reviewed, particularly those used in Allen and Brooks (1991) and our PET experiment, the retrieved exemplar may well have been explicitly represented (subjects could readily describe them verbally, for example).

Lastly, there are important categorization distinctions that pertain more directly to representations than processes. The best known of these distinctions come from neurological patients who have a category-specific deficit. Thus, a patient might have an inability to categorize animals while being relatively normal in the ability to categorize artifacts (e.g. Damasio, 1990; Warrington and Shallice, 1984). The dominant account of this dissociation is that animals tend to be represented mainly by perceptual attributes, whereas artifacts are represented as much by functional as perceptual attributes (e.g. Farah and McClelland, 1991). Although this perceptual-functional distinction is intended at the level of representations (or concepts), it might imply a distinction at the level of categorization procedures as well. Different kinds of detectors might be needed to recognize perceptual vs. functional features, and different kinds of detectors might be the starting point for qualitatively different categorization procedures.

But this is getting far ahead of what we know. What we do seem to have evidence for is the claim that rule application and exemplar similarity are very different psychologically and neurologically.

Acknowledgements

Some of the research reported in this paper was supported in part by a grant from the Office of Naval Research and in part by a grant from the National Institute on Aging.

References

Ahn, W., Brewer, W.F., Mooney, R.J., 1992. Schema acquisition from a single example. Journal of Experimental Psychology: Learning, Memory and Cognition 18, 391–412.

Ahn, W., Medin, D.L., 1992. A two-stage model of category construction. Cognitive Science 16 (1), 81–121.

Allen, S.W., Brooks, L.R., 1991. Specializing the operation of an explicit rule. Journal of Experimental Psychology: General 120 (1), 3–19.

Armstrong, S.L., Gleitman, L.R., Gleitman, H., 1983. What some concepts might not be. Cognition 13 (3), 263–308.

Awh, E., Smith, E.E., Jonides, J., 1995. Human rehearsal processes and the frontal lobes: PET Evidence. In: Grafman, J., Holyoak, K.J., Boller, F. (Eds.), Structure and Functions of Human Prefrontal Cortex. New York, NY: The New York Academy of Sciences, Volume 769.

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 (3), 278–287.

Bruner, J.S., Goodnow, J., Austin, G.A. (1956). A study of thinking, Wiley, New York, NY.

Bushnell, M.C., Goldberg, M.E., Robinson, D.L., 1981. Behavioral evidence of visual response in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. Journal of Neurophysiology 46 (4), 755–772.

Cohen, J.D., Forman, S.D., Braver, T.S., Casey, B.J., Servan-Schrieber, D., Noll, D.C., 1994. Activation of prefrontal cortex in a non-spatial working memory task with functional MRI. Human Brain Mapping 1, 293–304.

Damasio, A.R., 1990. Category-related recognition defects as a clue to the neural substrates of knowledge. Trends in Neuroscience 13 (3), 95–98.

D'Esposito, M., Detre, J., Alsop, D.C., Shin, R.K., Atlas, S., Grossman, M., 1995. The neural basis of the central executive system of working memory. Nature 378 (16), 279–281.

Diamond, A., 1988. The development and neural bases of memory functions as indexed by the AB and delayed response tasks in human infants and infant monkeys. In: Diamond, A. (Ed.), The Development and Neural Bases of Higher Cognitive Functions, The New York Academy of Sciences, New York, NY.

Estes, W.K., 1994. Classification and cognition, Oxford University Press, New York, NY.

Farah, M.J., 1990. Visual agnosia: disorders of object recognition and what they tell us about normal vision, MIT Press, Cambridge, MA.

Farah, M.J., McClelland, J.L., 1991. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. Journal of Experimental Psychology: General 120 (4), 339–357.

Fuster, J.M., 1995. Memory in the Cerebral Cortex, MIT Press Cambridge, MA.

Hampton, J.A., 1995. Testing the prototype theory of concepts. Journal of Memory and Language 34 (5), 686–708.

Hull, C.L., 1920. Quantitative aspects of the evolution of concepts, an experimental study. Princeton, NJ: Psychological review company.

Jaeger, J.J., Lockwood, A.H., Kemmerer, D.L., Van Valin, R.D. Jr., Murphy, B.W., Khalak, H.G., 1996. A positron emission tomography study of regular and irregular verb morphology in English. Language 72 (3), 451–496.

Kahneman, D., Miller, D.T., 1986. Norm theory: comparing reality to its alternatives. Psychological Review 93 (2), 136–153.

Keil, F.C., Batterman, N., 1984. A characteristic-to-defining shift in the development of word meaning. Journal of Verbal Learning and Verbal Behavior 23 (2), 221–236.
Keil, F.C., 1989. Concepts, kinds, and cognitive development, MIT Press, Cambridge, MA.
Kolodny, J.A., 1994. Memory processes in classification learning: An investigation of amnesic performance in categorization of dot patterns and artistic styles. Psychological Science 5 (3), 164–169.
Kemler-Nelson, D., 1984. The effect of intention on what concepts are acquired. Journal of Verbal Learning and Verbal Behavior 23, 734–759.
Knowlton, B.J., Ramus, S.J., Squire, L.R., 1992. Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances. Psychological Science 3, 172–179.
Kosslyn, S.M., Alpert, N.M., Thompson, W.L., Maljkovic, V., Weise, S.B., Chabris, C.F., Hamilton, S.E., Raunch, S., Buonanno, F.S., 1993. Visual mental imagery activates topographically organized visual cortex: PET investigations. Journal of Cognitive Neuroscience 5 (3), 263–287.
Kruschke, J.K., 1992. ALCOVE: an exemplar-based connectionist model of category learning. Psychological Review 99 (1), 22–44.
Luria, A.R., 1969. Frontal lobe syndromes. In: Vinken, P.J., Bruyn, G.W. (Eds.), Handbook of Clinical Neuropsychology (Vol. 2), North Holland. Amsterdam, Holland.
Milner, B., 1964. Some effects of frontal lobectomy in man. In: Warren, J.M., Akert, K. (Eds.), The frontal granular cortex and behavior, McGraw-Hill, New York, NY.
Medin, D.L., Schaffer, M.M., 1978. Context theory of classification learning. Psychological Review 85 (3), 207–238.
Medin, D.L., Wattenmaker, W.D., Hampson, S.E., 1987. Family resemblance, conceptual cohesiveness, and category construction. Cognitive Psychology 19 (2), 242–279.
Murphy, G.L., Medin, D.L., 1985. The role of theories in conceptual coherence. Psychological Review 92 (3), 289–316.
Nosofsky, R.M., 1986. Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General 115 (1), 39–57.
Nosofsky, R.M., 1992. Exemplars, prototypes, and similarity rules. In: Estes, W.K. (Ed.), From Learning Theory To Connectionist Theory: Essays in Honor of William K. Estes. Hillsdale, NJ: Lawrence Erlbaum.
Nosofsky, R.M., Palmeri, T.J., McKinley, S.C., 1994. Rule-plus-exception model of classification learning. Psychological Review 101 (1), 53–79.
Osherson, D., Smith, E.E., 1997. On typicality and vagueness. Cognition 64, 189–206.
Petersen, S.E., Fox, P.T., Posner, M.I., Mintum, M., Raichle, M.E., 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331, 585–589.
Pinker, S., 1991. Rules of language. Science 253 (5019), 530–535.
Posner, M.I., Petersen, S.E., Fox, P.T., Raichle, M.E., 1988. Localization of cognitive functions in the human brain. Science 240, 1627–1631.
Regehr, G., Brooks, L.R., 1993. Perceptual manifestations of an analytic structure: the priority of holistic individuation. Journal of Experimental Psychology: General 122 (1), 92–114.
Rips, L.J., 1989. Similarity, typicality, and categorization. In: Vosniadou, S., Ortony, A. (Eds.), Similarity and Analogical Reasoning, Cambridge University Press, Cambridge.
Rosch, E., 1973. On the internal structure of perceptual and semantic categories. In: Moore, T.E. (Ed.), Cognitive development and the acquisition of language, Academic Press, New York, NY.
Rubenstein, J., Evans, J.E., Meyer, D.E., 1994. Task switching in patients with prefrontal cortex damage. Paper presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.
Shanks, D.R., St. John, M.F., 1994. Characteristics of dissociable human learning systems. Behavioral and Brain Sciences 17, 367–447.
Shiffrin, R.M., Schneider, W., 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review 84 (2), 127–190.
Shimamura, A.P., 1994. Memory and frontal lobe function. In: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences, MIT Press, Cambridge, MA.

Smith, E.E., Jonides, J., Koeppe, R.A., 1996. Dissociating verbal and spatial working memory using PET. Cerebral Cortex 6, 11–20.

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

Smith, E.E., Medin, D.L., 1981. Categories and concepts, Harvard University Press, Cambridge, MA.

Smith, E.E., Patalano, A.L., Jonides, J., Koeppe, R.A. (Submitted). Pet evidence for two kinds of categorization.

Smith, E.E., Shoben, E.J., Rips, L.J., 1974. Structure and process in semantic memory: a featural model for semantic decisions. Psychological Review 81 (3), 214–241.

Smith, E.E., Sloman, S.A., 1994. Similarity- versus rule-based categorization. Memory and Cognition 22 (4), 377–386.

Smith, J.D., Kemler, D.G., 1984. Overall similarity in adults' classification: the child in all of us. Journal of Experimental Psychology: General 113 (1), 137–159.

Smith, J.D., Shapiro, J.H., 1989. The occurrence of holistic categorization. Journal of Memory and Language 28 (4), 386–399.

Squire, L.R., 1992. Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. Psychological Review 99 (2), 195–231.

Squire, L.R., Knowlton, B.J., 1995. Learning about categories in the absence of memory . Proceedings of the National Academy of Science, USA 92, 12470–12474.

Ullman, M., Corkin, S., Coppola, M., Hickok, G., Growdon, J.H., Koroshetz, W.J., Pinker, S., 1997. A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. Journal of Cognitive Neuroscience 9, 266–276.

Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. Brain 107, 829–853.