Journal of Experimental Psychology:
Human Learning and Memory
1981, Vol 7, No. 4, 241-253

0096-1515/81/0704-0241$00.75

Strategies and Classification Learning

Douglas L. Medin
University of Illinois at Urbana-Champaign

Edward E. Smith
Stanford University

How do strategies affect the learning of categories that lack necessary and sufficient attributes? The usual answer is that different strategies correspond to different models. In this article we provide evidence for an alternative view—Strategy variations induced by instructions affect only the amount of information represented about attributes, not the process operating on these representations. The experiment required subjects to classify schematic faces into two categories. Three groups of subjects worked with different sets of instructions: roughly, form a prototype of each category, learn each category as a rule-plus-exception, or standard neutral instructions. In addition to learning the faces (Phase 1), subjects were given transfer tests on learned and novel faces (Phase 2) and speeded categorization tests on learned faces (Phase 3). There were performance differences in all three phases due to instructions, but these results were readily accounted for by specific changes in the representations posited by the context model of Medin and Schaffer; that is, strategies seemed to affect only the amount of information stored about each exemplar's attributes.

The recent upsurge of interest in natural categories such as bird, tree, and fruit has been accompanied by parallel investigations of representations and processing of artificial categories. In this article we are primarily concerned with the role of strategies in learning the attribute structure of artificial categories. But since the effects of strategies can best be understood in terms of specific categorization models, we first provide a brief overview of models and then take up the strategy issue.

Category Learning Models

One idea growing out of research with artificial categories is that based on experience with exemplars, people abstract some measure of the central tendency of a category and base their categorical judgments on this central tendency, or *prototype* (e.g., Posner & Keele, 1968). A contrasting view posits that when an item is presented to be classified, it acts as a retrieval cue to access information associated with similar stored exemplars, and this specific exemplar information is the basis for category judgments (Medin & Schaffer, 1978). According to the latter idea, some animal might be categorized as a rodent not on the basis of a comparison to a rodent prototype, but because that animal has similar attributes to a rabbit and the categorizer thinks that rabbits are rodents. A specific proposal embodying this idea, known as the context model, will be considered in detail shortly.

Since both prototype and exemplar-based models can account for many phenomena, it is difficult to generate differential predictions (see, e.g., Hintzman & Ludlam, 1980). In one attempt to do so, Medin and Schaffer (1978) contrasted the predictions of what they called *independent-* and *interactive-cue* theories, with prototype models being one type of independent-cue theory. Independent-cue theories assume that the information entering into category judgments (overall similarity, distance, or validity) can be derived from an additive combination of the information from component attributes (Franks & Bransford, 1971; Reed, 1972). In other words, the more characteristic attributes an exemplar has, the easier it should

This research was supported in part by National Institute of Mental Health Grants MH 32370 and MH 19705 and by the National Institute of Education Contract HEW-NIE-C-400-76-0116.

Edward E. Smith is now at Bolt Beranek & Newman Inc., Cambridge, Massachusetts.

Requests for reprints should be sent to Douglas L. Medin, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

be to learn and classify. Interactive-cue theories reject such additivity. Thus in Medin and Schaffer's (1978) model, which is an interactive-cue model, the various attribute values comprising exemplars are combined in a multiplicative manner to determine the overall similarity of two exemplars.

The multiplicative rule has the implication that an exemplar may be classified more efficiently if it is highly similar to one instance and dissimilar to a second than if it has medium similarity to two instances of a category. Hence the context model predicts that categorization performance will vary with the number of stored exemplars similar to the test item. Independent-cue models are insensitive to such density effects. In a series of four experiments, Medin and Schaffer (1978) obtained clear support for the context model. Data from original learning, transfer, and speeded classification were in each case more in line with the context model than with a generalized independent-cue model. In addition, a mathematical version of the context model gave an excellent quantitative account of classification performance on transfer tests involving new and old instances.

The Strategy Issue

One response to the above results is to question their generality, particularly with respect to the issue of strategies. One could argue that there was something about the Medin and Schaffer (1978) items, or some detail of the experimental situation, that discouraged people from developing the type of category representation appropriate to independent-cue theories, like a prototype model. If people had been instructed, say, to form a prototype, the results might well have been different.

The importance of strategies in classification learning seems undeniable, and there are important issues concerning how strategies ought to be treated by a theory. One point of view is that strategies modify both representations and processes. According to this idea, to understand categorization one needs (a) a list of the possible strategies that might be used in the task, (b) a separate theory mapping each strategy on to performance, and (c) a higher level theory specifying the factors governing strategy selection. Under this view current theories of categorization are essentially alternative procedures, each of which can be evidenced when their eliciting factors are operative. In this sense all models are correct (and all incorrect) at least some of the time.

An alternative view is that strategies induced by instructions alter the underlying representation in particular ways while leaving unchanged the processes operating on these representations. For example, a key assumption of independent-cue models is that judgments are based on a weighted, additive combination of information from component attributes. Strategies might alter the weights attached to different attributes, but judgments could still be based on an additive and independent combination of information. In other words, strategies would influence the parameters in the model but leave the basic model intact. The same possibility holds for interactive-cue models, such as the context model. One could hold that strategy variations modify category representations in essentially quantitative ways while leaving the retrieval process implied by the context model unchanged. The multiplicative rule for similarity relationships would still hold, but the similarity parameters associated with component attributes might well differ for different strategies. The goal of the present investigation was to provide evidence bearing on this alternative view of strategies.

Overview of the Experiments

In the present experiments we attempted to induce strategy variations by means of instructions. The task involved two fuzzy categories; that is, no individual cue was perfectly valid and associated with members of one category but not the other. In one condition people were asked to use a rule-plus-exception strategy; in a second people were asked to learn the central tendency, or prototypes, of the categories; in the third people were given no special instructions. Our aim was to see if the model could handle learning and transfer data from these three distinct conditions solely in terms of differences in the similarity parameters associated with the attributes.

By imposing some instructional control over the strategies that people employed, we hoped to obtain a clearer picture of the effects of strategies on performance as well as an indication of what aspects of the data were invariant over strategies. The instructional variations were also designed to put particular models, for example, prototype models, in the best light by asking people to do what the models imply that they normally do.

The basic design of the experiments is shown in Table 1. The items were Brunswik faces varying in eye height (EH), eye separation (ES), nose length (NL), and mouth height (MH). There were two possible values for each attribute, which are represented in the table in terms of a binary notation. For example, the value 1 on the attribute of nose length might correspond to a long nose, the value 0 to a short nose. Categories A and B differ with respect to what is generally true. That is, on each attribute Category A exemplars tend to have the value 1 and Category B exemplars the value 0, although there is at least one exception for each attribute in each category. This distribution of attributes corresponds to that used by Medin and Schaffer (1978) in their Experiments 2 and 3; as they note, the two categories are separable by a linear discriminant function as required by independent-cue models (see, e.g., Reed, 1972).

Table 1
Attribute Structure of Categories Used in the Experiment

Face no.	Attribute value: EH	ES	NL	MH
Training items				
A exemplars				
4	1	1	1	0
7	1	0	1	0
15	1	0	1	1
13	1	1	0	1
5	0	1	1	1
B exemplars				
12	1	1	0	0
2	0	1	1	0
14	0	0	0	1
10	0	0	0	0
New transfer items				
1	1	0	0	1
3	1	0	0	0
6	1	1	1	1
8	0	0	1	0
9	0	1	0	1
11	0	0	1	1
16	0	1	0	0

Note. EH = eye height; ES = eye separation; NL = nose length, MH = mouth height. See the text for explanation of binary notation.

Although the models will be evaluated in terms of their ability to account for the entire pattern of data, a good guidepost for distinguishing the context model from independent-cue models is the comparison of Face 4 and Face 7 (see Table 1). Since the central tendency, or modal prototype, for Category A is 1111, Face 4 must be at least as close to the prototype as Face 7 regardless of how the attributes are weighted.[1] Thus all independent-cue models predict that Face 4 will be easier to learn and more accurately classified than Face 7 because for the only dimension where the two differ, Face 4 has the typical or characteristic value and Face 7 the atypical value. In contrast, interactive-cue models in general and the context model in particular predict Face 7 should be easier because the number of highly similar patterns is the most important factor in performance. Although one would not want to assume all dimensions are equally salient, for convenience we shall call two faces highly similar if they differ in value along only one dimension. Face 7 is highly similar (differs on only one attribute) to two other faces in Category A (4 and 15) but is not highly similar to any face in Category B. Face 4, on the other hand, is highly similar to one face in Category A (7) and to two in Category B (2 and 12); it should be more difficult to classify. This prediction is not entirely parameter free, but it holds over a very large

[1] Although we describe the prototype in terms of modal values, a prototype could as readily be based on mean values. This distinction would not lead to differential predictions in our experiments, as both a modal prototype and a mean prototype represent special cases of the general independent-cue model.

range of values, and values that would alter that prediction would produce other testable distinctions between the context and independent-cue models.

A Note on Stimuli and Category Structure in Relation to the Models

The generality of any results from the particular stimuli and category structure shown in Table 1 are obviously limited. In what follows we provide some rationale for our particular choice of structure and stimuli to clear up some common misconceptions about the models.

Category structure. The categories were set up to allow for a contrast between the context model and independent-cue models without violating any major constraint on structure implied by these models. The only constraint associated with the context model is that categorization will be easier to the extent that within-category similarity is maximized and between-category similarity is minimized. Independent-cue models require that categories be linearly separable if perfect categorization is to occur. That is, there must be some weighted additive combination of attribute values that puts all A exemplars into Category A and all B exemplars into Category B. One way to see that the stimuli in Table 1 satisfy that constraint is to note that all exemplars could be correctly classified by looking at eye height, nose length, and mouth height, determining whether the values were typical of category A or B, and then assigning the face to Category A if two or more of the values were typical for Category A and to Category B if two or more of the values were typical for B. When categories are linearly separable, it is more difficult to distinguish the predictions of the contending models, but as we have seen, the structure in Table 1 does permit some contrast.

Stimuli. One reason for using Brunswik faces was that Reed (1972) used them in studies that were taken as providing support for a specific independent-cue model (prototype theory). These experiments predated the context theory, so it is difficult to judge how it would have fared.

A second major reason for using Brunswik faces is that they are highly confusable. Though the individual values are distinctive, individual faces differ from each other only in the particular combination of attribute values they possess. This might appear to bias the experiment against the context model inasmuch as it assumes that performance is based on the retrieval of specific item information; however, the context model does not specifically assume that a distinct or distinctly accessible representation is developed for each individual stimulus. In the next paragraphs we amplify this point. (For additional details, see Medin & Schaffer, 1978, pp. 210–212.)

In the context model the parameter reflecting the similarity of two values on an attribute is assumed to be less when that attribute is attended to, or forms part of a hypothesis, than when it is not so salient. If instructions encourage the belief that there is only one critical attribute, many attributes of the item will not be encoded in any detail at all, and the resulting representations will not be sufficient to distinguish the individual exemplars (e.g., Bourne & O'Banion, 1969; Calfee, 1969). In other words, it is not necessarily assumed that a distinct representation is set up for each individual exemplar.

Consider a highly simplified classification task involving two Category A patterns (A_1 = 1110, A_2 = 1010) and two Category B patterns (B_1 = 0001, B_2 = 0010). Suppose that a person in the experiment has selectively attended (perhaps tested hypotheses about) the second and third dimensions, so less information has been stored about the first and fourth dimensions. The subject's representation of exemplar information might be something like this:

$$?11? - A(A_1) \quad 000? - B(B_1)$$

$$?010 - A(A_2) \quad ?010 - B(B_2)$$

where the question marks indicate that information that would differentiate Value 1 and Value 0 on that dimension was not stored (or cannot be accessed). Note that this representation is not sufficient to produce perfect performance because the representations associated with A_2 and B_2 cannot be distinguished. When B_2 is presented for a test, the representation associated with

A_2 should be as likely to be accessed as the representation associated with B_2. If a new pattern B_3 = 0000 is presented, it should be correctly classified because it would most likely access the representation associated with B_1. Note further that on a new–old recognition test, B_3 would very likely be recognized falsely as old for the same reason. Thus depending on the completeness of the exemplar information and the nature of the probes, new–old recognition could actually be at chance and classification could be based on exemplar representations and be relatively accurate.[2]

By using highly confusable stimuli, we aimed to assure that when a training stimulus was presented, it would not automatically retrieve its own representation. As we have seen the context model does not assume that exemplars get perfectly coded nor that a probe invariably accesses the corresponding representation in memory. Even when the correct category assignment has been attached to each of the individual training stimuli, classification performance may not be perfect because the presentation of a training stimulus would not invariably lead to accessing its associated representation in memory.

A secondary aim of using confusable stimuli was to make clearer the nature of the contrast between independent-cue models and the context model. Although the context model assumes that performance is based on retrieval of exemplar information and certain versions of the general independent-cue model assume that performance is based on an abstract prototype, the fundamental contrast is not between exemplar and prototype models but, rather, between independent-cue and interactive-cue models. For example, one independent-cue model, the average distance model (e.g., Reed, 1972), is an exemplar model that assumes that representations associated with every training stimulus are retrieved when a probe is presented and that the probe is assigned to the category whose members have the greater average similarity to (lesser average distance from) the probe item. As long as binary-valued dimensions are used, as in the present experiments, the predictions of an average distance model cannot be differentiated from predictions of a prototype model. It is also true that the context model and some (but not all) other interactive-cue models assume that categorization can involve a considerable amount of abstraction—The key difference is that interactive-cue models do *not* assume that this abstracted information is confined to an additive and independent sum of components.

Finally, one should note that more general forms of independent-cue models allow for training stimuli to be classified on the basis of specific information concerning that particular item. To the extent that this occurs, it would be more difficult to distinguish predictions of the contending theories, at least with respect to classification of training stimuli. Therefore, to minimize information specific to individual faces, the stimuli differed from each other only in their combination of values.

Method

Subjects

Ninety-six volunteers were solicited through ads in local newspapers. The subjects, men and women ranging in age from 17 to 30 years, were paid $2.50 for the experimental session. Thirty-two people were assigned to each of the three instructional conditions.

Stimuli

The stimuli were Brunswik faces displayed on an approximately 27 cm × 34 cm visual display screen (Digital Equipment Corp. VR-17 cathode ray tube screen) linked to a PDP-11 computer. The face outlines were 13.5 cm × 11.5 cm and centered on the screen. The faces

[2] The above offers one interpretation of the context model that makes it qualitatively consistent with new–old recognition being poor yet classification being exemplar-based and relatively accurate. Other interpretations of the context model are also possible here. Shiffrin (Note 1) has suggested that there is some probability on any training trial that one or more attribute values of the pattern may be encoded erroneously. Thus though the most likely outcome of a training trial is that all encoded values are correct, the second most likely outcome is that one encoded value is in error, the third most likely outcome is that two encoded values are in error, and so forth. The upshot is that training results in many exemplars of various sorts being stored in memory, with the likely majority being the exemplars actually presented or single-value perturbations of them. This majority underlies relatively accurate exemplar-based classification, and the sheer number of stored exemplars would be responsible for poor new–old recognition.

differed in nose length, mouth height, eye separation, and eye height—the same attributes that Reed (1972) varied in his studies of category learning. The nose was either a 1.5-cm or a 3.0-cm vertical line centered within the face outline. The mouth was a 4.0-cm horizontal line, which was either 1.5 cm or 3.0 cm from the chin line. The eyes were 1 cm × 2.5 cm and were separated by either 1.5 cm or 3.5 cm (measuring from inner edges). Finally, the eyes were either 2.5 cm or 5 cm from the top of the face outline (measured to the top edge of the eye). The two possible values on each of the four attributes were combined to produce 16 distinct faces. The categories were constructed in accordance with the design shown in Table 1.

All subjects were presented the same faces, but the particular assignment of individual faces to the abstract notation in Table 1 varied across subjects. For example, for one subject 1001 might refer to a face with eyes up and far apart, a long nose, and a low mouth; for another subject 1001 might refer to a face with eyes down and close together, a long nose, and a low mouth; and so on. Overall, each face was assigned to a given abstract notation exactly twice for each instructional condition, once when Faces 4, 6, 7, 13, and 15 were associated with Category A and once when they were associated with Category B. Hence the assignment of faces to conditions, instructions, and category labels was completely counterbalanced.

General Procedures

The procedure had three phases: original learning, transfer, and speeded classification. Each phase is described below.

Initial learning. This phase consisted of up to 32 runs through the set of nine training faces (see Table 1), with a learning criterion of one errorless run. The trial sequence went as follows: (a) A face appeared on the screen. (b) To indicate their categorization, subjects pressed either the button marked *A* or the button marked *B*, which occupied the lower left and right corners, respectively, of a 4 × 4 button response box. (3) The face remained on the screen for an additional 2 sec while feedback was displayed below the face. (d) A 1-sec intertrial interval ensued.

The first part of the instructions was the same for all three instructional conditions. Subjects were told that they would see faces differing only in nose length, mouth height, distance between the eyes, and height of the eyes, and that their task was to learn to correctly classify the faces into Category A or Category B. They were further told that each facial feature had some information value for category membership, but that none was a perfectly reliable indicator of category membership.

The general procedure was then described. Subjects were told that they would be given immediate feedback about the correctness of their categorization responses. Subjects were then given the specific instructions designed to induce strategy differences. One group was given standard instructions not specifying any particular strategy[3]; a second group was told to use a rule-plus-exception strategy; the third group was asked to learn the central tendency (or prototype) for each category. Details are provided in *Instructions for Initial Learning.* Note that none of the groups were given instructions to use exemplars as a basis of performance, though this is the process specified by the context model.

Transfer. Transfer tests immediately followed initial training. Subjects were instructed they would see the old faces mixed in with some new faces that were very similar to the old. As each face appeared they were asked to check to see whether it was a new face and, if so, to press the button marked *N* for new. Then regardless of whether the face was old or new, they were to decide whether the face belonged in Category A or Category B, based on what they had learned before. Finally, they were asked to indicate how confident they were concerning which category the face belonged to by pressing the *Guess* button, the *Think so* button, or the *Sure* button. Subjects were then given two runs through all 16 faces, using different randomizations of the faces in the two runs. Each face remained on the screen until the confidence judgment was given. The intertrial interval was as before, but no feedback concerning either recognition or classification was given.

Speeded classification. After the transfer tests subjects were given an additional 16 runs through the nine training faces. Presentation and feedback were exactly as in initial learning; the only difference was that subjects were now told that their latencies were being measured and they were to respond as fast as they could without making errors. Subjects in the various instructional conditions were asked to use the strategy they had employed before.

Instructions for Initial Learning

Standard instructions. These instructions were not designed to focus subjects on any particular strategy. They read as follows:

> At first you will have to guess because I haven't given you any information about which category a particular face falls into. Each face will be presented repeatedly during the learning phase of the experiment and by paying attention to the feedback, eventually you can be able to correctly assign each face to its appropriate category.

Rule-plus-exception instructions. In addition to the above standard instructions, subjects were told:

> We want you to use a particular strategy to learn to classify the faces. You might call it a "rule-plus-exception strategy". First pay attention to nose length and learn which category most short-nosed faces fall into and which category is correct for most long-nosed faces. You will find that one short-nosed face and one

[3] This instruction group is actually Experiment 3 of Medin and Schaffer (1978). The full description is reported here to facilitate cross-referencing and because previously unreported data from this group are included in the present article.

long-nosed face are exceptions to the rule. Memorize these faces. When you have mastered the task, you will be doing something like looking to see if the face is one of the exceptions; if so, make the correct response; if not, apply the rule for short and long noses. Usually the task would be quite difficult and fewer than half the people who try it would be able to learn it, but by using the rule-plus-exception strategy and focusing on nose length, you should be able to readily learn which category is correct for each face.

Prototype instructions. Since a prototype might be based on either mean or modal values, the instructions were designed to be neutral on this point. In addition to the standard instructions, subjects were told:

> There are probably many strategies you could use to learn which category each face belongs to but we want you to focus on one particular one. As the faces appear, we want you to form a general impression of what "A" faces on the average look like and what "B" faces on the average look like. At the end of the learning phase of this experiment, I'll ask you whether the faces in one category generally had short or long noses, low or high mouths, up or down eyes, or close or far-apart eyes. As you develop a general impression of what "A" faces on the average look like and what "B" faces on the average look like, we want you to use these general impressions to help you classify the faces.

General Results

Initial Learning

As anticipated the initial learning task proved difficult, with fewer than half of the subjects meeting the criterion of one errorless run. Overall, 14 of 32 people met the criterion in the standard and rule-plus-exception conditions, but only 8 of 32 reached criterion in the prototype condition. All subjects, however, improved with practice. If performance were completely at chance, subjects should have averaged 16 errors per face during learning (each face was presented 32 times); instead average errors per face were 8.0 under standard instructions, 6.4 under rule-plus-exception instructions, and 9.2 under prototype instructions.

Table 2 shows the distribution of errors across faces for the three instructional conditions. The rule-plus-exception instructions led to the best overall performance; the prototype instructions resulted in the poorest overall performance. There were also considerable differences across instructional conditions in the relative difficulty of individual faces. Face 12, for example, was associated with an average of 17.4 errors under prototype instructions but only 6.3 errors under rule-plus-exception instructions. Statistical tests confirm the reliability of these differences. An analysis of variance was conducted on errors during learning using the factors of instructional conditions, randomizations, and faces. Significant effects were obtained for the main effects of instructions, $F(2, 48) = 4.44$, $MS_e = .93$, $p < .02$; faces, $F(8, 384) = 55.96$, $MS_e = .69$, $p < .0001$; and the Instructions $\times$ Faces interaction, $F(16, 384) = 3.96$, $MS_e = .69$, $p < .001$. In short, different instructions produced differences in learning that interacted with particular faces.

There were also some constants across instructions. In particular, the theoretically important comparison of Faces 4 and 7 showed a small but reliable advantage for Face 7 in all conditions, $t(95) = 2.87$, $p < .01$, consistent with the context model but not with independent-cue models. Also the faces that served as exceptions in the rule-plus-exception condition (Faces 2 and 13) were among the most difficult to master in all conditions.

Transfer Tests

Recognition. Recognition performance was very poor and there were only minor

Table 2
Mean Number of Errors for Each Face During Initial Learning as a Function of Instructions

Face Number	Instruction: Standard	Rule-plus-exception	Prototype
4	4.5	3.9	7.7
5	8.2	5.9	9.2
7	4.2	3.3	6.7
13	11.9	10.7[a]	13.7
15	2.8	2.8	4.9
2	12.9	13.8[a]	10.3
10	4.4	3.8	4.2
12	15.2	6.3	17.4
14	6.6	6.8	8.7
M	8.0	6.4	9.2

[a] Face was an exception in the rule-plus-exception condition.

differences across instruction conditions. The probability of saying new to an old face was .19, .22, and .13 in the respective standard, exception, and prototype instructional conditions, but the probability of saying new to a new face was only .23, .26, and .19 in the respective conditions. Overall, 59 subjects were more likely to say new to a new face than to an old face, 9 showed no difference, and 28 showed the opposite trend. Across the 96 subjects new–old recognition was barely above chance $\chi^2(1) = 6.44$, $p < .05$. As noted in the introduction, this result is consistent with both classes of models under consideration.

Categorization during transfer. Categorization accuracy for the nine old and seven new faces for the various instructional conditions is shown in the first columns of Tables 3, 4, and 5. (The remaining columns of these tables contain theoretical predictions that will shortly be discussed.) Again prototype instructions yielded the poorest overall performance. Of course this may merely reflect the fact that this group mastered less of the categorical structure during the initial learning phase. Of somewhat greater interest are the differences between the instructional conditions in the relative difficulty of individual new faces. For example, with rule-plus-exception instructions, Face 1 was categorized as an A 45% of the time and Face 3 as a B 80% of the time; in contrast, with prototype instructions Face 1 was called an A 73% of the time and Face 3 a B 35% of the time. But importantly, although the differences were small, Face 4 was never categorized more accurately than Face 7 under any instructions. Moreover, the largest difference favoring Face 7 occurred under the prototype instructions, which should have been maximally favorable to the independent-cue model. Overall, Face 7 was classified significantly more accurately than Face 4, $t(95) = 2.89$, $p < .02$.

An analysis of variance of responses to the old faces revealed significant effects of instructions, $F(2, 48) = 5.32$, $MS_e = .93$, $p < .01$, and faces, $F(8, 384) = 17.56$, $MS_e = .84$, $p < .001$. A similar analysis for the new faces produced significant effects for faces, $F(6, 288) = 38.22$, $MS_e = 1.00$, $p < .0001$, and for the Instructions × Faces interaction, $F(90, 288) = 8.37$, $MS_e = 1.00$, $p < .0001$. (Detailed accounts of these results will be offered in the Theoretical Analysis section.)

Table 3
Observed and Predicted Proportions of Correct Categorizations for Each Face During Transfer: Standard Instructions

Face number and category label	Proportions		
	Observed	Predicted context model	Predicted independent-cue model
	Old faces		
4A	.97	.94	.95
7A	.97	.99	.92
15A	.92	1.00	.96
13A	.81	.72	.79
5A	.72	.71	.71
12B	.67	.68	.67
2B	.72	.71	.76
14B	.97	1.00	.95
10B	.95	1.00	1.00
M	.89		
	New faces		
1A	.72	.78	.59
6A	.98	.95	1.00
9A	.27	.30	.14
11A	.39	.47	.43
3B	.44	.45	.49
8B	.77	.78	.65
16B	.91	.88	.94

Speeded Classification

Table 6 summarizes the data from the speeded-classification phase of the experiment. Average correct reaction times are given for each face for each instructional condition, with corresponding error rates. The pattern of results is by now a familiar one. There are instructional differences in the relative difficulty of faces—for example, Faces 2 and 13, the two exceptions, are the most difficult only in the rule-plus-exception condition. And consistent with the context model, once more Face 7 resulted in better performance—faster reaction times and fewer errors—than Face 4 in all instructional conditions. Again, this effect was modest but consistent.

Table 4
Observed and Predicted Proportions of Correct Categorizations for Each Face During Transfer: Rule-Plus-Exception Instructions

Face number and category label	Proportions: Observed	Predicted context model	Predicted independent-cue model
	Old faces		
4A	.89	.91	.92
7A	.94	.97	.91
15A	.94	.99	.98
13A	.72	.67	.68
5A	.78	.74	.84
12B	.73	.72	.82
2B	.70	.61	.66
14B	.91	.96	.92
10B	.95	.98	1.00
M	.84		
	New faces		
1A	.45	.50	.41
6A	.88	.94	1.00
9A	.08	.20	.16
11A	.75	.79	.69
3B	.80	.80	.72
8B	.42	.48	.44
16B	.88	.92	.97

Table 5
Observed and Predicted Proportions of Correct Categorizations for Each Face During Transfer: Prototype Instructions

Face number and category label	Proportions: Observed	Predicted context model	Predicted independent-cue model
	Old faces		
4A	.77	.83	.84
7A	.97	.89	.92
15A	.98	.93	.98
13A	.70	.74	.77
5A	.60	.57	.52
12B	.45	.47	.51
2B	.72	.70	.75
14B	.83	.85	.84
10B	.87	.91	1.00
M	.79		
	New faces		
1A	.73	.77	.72
6A	.87	.89	1.00
9A	.28	.28	.20
11A	.52	.46	.44
3B	.35	.40	.46
8B	.78	.74	.74
16B	.88	.85	.98

Statistical tests confirm the above impressions. Separate analyses of variance were conducted on reaction times and errors. For the reaction times, there were significant effects of instructions, $F(2, 48) = 13.72$, $MS_e = 2.21$, $p < .001$; faces, $F(8, 384) = 17.93$, $MS_e = .20$, $p < .0001$; and the Instructions × Faces interaction, $F(16, 384) = 3.09$, $MS_e = .20$, $p < .0001$. For errors, there were significant effects of faces, $F(8, 384) = 38.06$, $MS_e = .79$, $p < .0001$, and of the Faces × Instructions interaction, $F(16, 384) = 2.63$, $MS_e = .79$, $p < .001$. Also, the three-way Instructions × Faces × Randomizations interaction was marginally significant, $F(120, 384) = 1.30$, $MS_e = .79$, $p < .05$. Planned t tests on Faces 4 and 7 indicated that the latter produced fewer errors, $t(95) = 3.18$, $p < .01$, and faster responding, $t(95) = 1.91$, $p < .06$.

Theoretical Analysis

Although the guideline comparison of Face 4 with Face 7 uniformly favored the context model, it is also important to see how the contending models fit the transfer data quantitatively. For the context model pre-

Table 6
Mean Correct Reaction Times (RT; in msec) for Each Old Face During Speeded Classification as a Function of Instructions

Stimulus number	Standard RT	Standard ER	Rule-plus-exception RT	Rule-plus-exception ER	Prototype RT	Prototype ER
4	1.11	.05	1.27	.03	1.92	.07
5	1.34	.14	1.61	.11	2.13	.18
7	1.08	.03	1.21	.01	1.69	.04
13	1.27	.09	1.87[a]	.15	2.12	.14
15	1.07	.02	1.31	.01	1.54	.04
2	1.30	.12	1.97[a]	.20	1.91	.12
10	1.08	.03	1.42	.02	1.64	.03
12	1.37	.19	1.58	.10	2.29	.16
14	1.13	.06	1.34	.04	1.85	.06
M	1.19	.08	1.51	.07	1.90	.09

Note. ER = error rate.
[a] Stimuli that were exceptions to the rule.

dictions concerning transfer can be analyzed in terms of which exemplars are likely to be retrieved when any particular face is presented as a probe. It is assumed that the probability of assigning a particular probe face to Category A(B) is equal to the sum of the similarities of that face to each of the stored exemplars of A(B), divided by the sum of the similarities of that face to each of the stored exemplars of both categories (Medin & Schaffer, 1978).[4] The similarity between a probe face and an exemplar is determined by a multiplicative combination rule. More precisely, the model has four parameters (each ranging between 0 and 1, with 1 designating maximum similarity) that correspond to the similarity parameters for the values of the four attributes used in this experiment; for example, the parameter for eye height specifies the similarity between the two different values of this attribute. And similarity between two items is determined by multiplying the relevant parameters.

Quantitative predictions of the general independent-cue model (including the prototype model as a special case) require some additional assumptions. Following Medin and Schaffer's (1978) treatment of independent-cue theory, we assume that associated with each attribute is some weight parameter reflecting the importance of that attribute in categorization. If W_{eh}, W_{es}, W_{nl}, and W_{mh} are the weight parameters associated with eye height, eye separation, nose length, and mouth height, respectively, then the probability that a particular face will be classified as an A is equal to the relative weight of values consistent with Category A. For example, the probability that a face with values 1010 would be called an A is $(W_{eh} + W_{nl})/(W_{eh} + W_{es} + W_{nl} + W_{mh})$. Since we are working with ratios, there are really only three independent paremeters, assuming the parameters sum to one. In addition, in the case of old faces, we will assume that transfer performance may be based on specific exemplar information. This probability is represented by a parameter S. Since 1010 is a face used in training (Face 7), the probability of its being classified as an A would be equal to $S + (1.0 - S) \times (W_{eh} + W_{nl})/(W_{eh} + W_{es} + W_{nl} + W_{mh})$.

Both the general independent-cue model and the context model thus have four free parameters to be used in fitting the data. These parameters were separately estimated (by minimizing least squares) for each of the three instructional conditions. The goal of both models is to describe the transfer data from the different instructional conditions in terms of only parameter changes. The predictions from two models are shown in Tables 3, 4, and 5. Both models do a fair job of capturing the main trends in the data, but the context model seems to provide better quantitative predictions. Some evidence for the latter's superiority is presented in Table 7. The table gives three measures of each model's goodness-of-fit—average absolute deviation, sum of squared deviations, and rank order correlation between predicted and observed classification accuracy—and all favor the context model.

To obtain a more precise comparison of how well the two models fit the data, chi-square tests were computed for the three experiments. In these tests all cases in which the expected number of responses was less than five were lumped into a single cell. For the independent-cue model, $\chi^2(31) = 194.4$, $p < .001$, which is highly significant; for the context model, $\chi^2(33) = 47.3$, $p \simeq .05$, which is not quite statistically significant. The difference in chi-square values provides an index of relative accuracy of the two models and this difference, $\chi^2(1) = 147.1$, is highly significant, ($p < .001$). (Technically this test requires that the component chi-squares have the same degrees of freedom, but the fact that there were fewer degrees of freedom associated with the independent-cue model should, if anything, favor this model.)

The parameter values associated with these predictions are shown in Table 8. The parameter constraints are fairly tight in that values more than a few percentage points

[4] As Medin and Schaffer (1978) noted, the idea is not that all stored patterns are accessed by each probe but, rather, that the similarity parameters determine which patterns are likely to be accessed by the probe. The particular ratio rule is already an approximation, since similarity parameters would be expected to differ for individual subjects. The best defense of the response rule is that it is a fair approximation and that it seems to work.

Table 7
Statistics for Evaluating Goodness of Fit of Models Applied to Transfer Categorization Data

Instructional condition	Average absolute deviation		Σ deviations squared		Rank-order correlations	
	Context	Independent cue	Context	Independent cue	Context	Independent cue
Standard	.036	.048	.032	.063	+.90	+.90
Rule-plus-exception	.048	.055	.049	.062	+.98	+.87
Prototype	.038	.072	.028	.114	+.96	+.88

away yield substantially poorer fits for both models. Consider first the parameters for the context model. Here the smaller the similarity parameter, the more salient the dimension. Note that in the rule-plus-exception condition, the similarity parameter for the dimension of nose length is 0, which is consistent with the instructions, making this attribute salient. (There is no parameter for specific item information, since performance is assumed to be exclusively based on exemplar retrieval.) Most important, note that the parameters for standard and rule-plus-exception instructions are generally smaller than those with prototype instructions. This suggests that subjects who had either standard or rule-plus-exception instructions were more likely to store distinctive information about the attributes of exemplars than subjects who had prototype instructions. And this accounts for why the former two instructional conditions led to better overall performance than did prototype instructions. Similarly, the changes in parameter values with instructions account for many of the interactions we obtained between faces and instructions. As one example, Face 12 was categorized more efficiently with rule-plus-exception than prototype instructions because (a) under the former instructions nose length is more salient and (b) correct categorization of this face hinges critically on the nose-length attribute (see Table 1).

For the independent-cue model, the greater the parameter value associated with an attribute, the greater its weight or salience. Therefore these values generally should be and are negatively correlated with the similarity parameters of the context model. The parameter values for specific exemplar information raise a problem for the independent-cue model. In two of the three instructional conditions this value is high (.42). This leads one to ask why old–new recognition was so poor, and why it was not poorer with prototype instructions than with the other two conditions. That is, if there is information that identifies specific exemplars during classification, then that same information should have mediated old–new recognition. The fact that it did not casts further doubt on the independent-cue model, at least the versions that include specific item information. We also tried fitting the data with a variation of the independent-cue model that did not include specific item information but assumed that there was some probability, p, that the summary representation (e.g, prototype) had not been developed, in which case it was assumed that subjects were forced to guess. This variation of the model produced markedly worse fits to the data mainly because performance on old face patterns was too good relative to performance on new faces, a result that the modified model cannot predict.

A Comment About Degree of Learning

The question arises about whether there were any systematic differences between learners and nonlearners, or differences as a function of stage of learning. For example, it is conceivable that the independent-cue model describes performance accurately only early in learning, whereas the context model might be more accurate at later stages of learning. The data give little support to this idea. The relative difficulty of Faces 4 and 7 did not show any noticeable interaction with practice, and Face 7 was easier than Face 4 for both learners (those who met the

Table 8
Best Fitting Parameter Values for the Two Models

Dimension	Eye height	Eye separation	Nose	Mouth	Specific item (S)
Standard instructions					
Context	.01	.11	.07	.40	
Independent cue	.50	.06	.36	.08	.42
Rule-plus-exception instructions					
Context	.15	.37	.00	.41	
Independent cue	.28	.03	.55	.13	.42
Prototype instructions					
Context	.05	.62	.37	.48	
Independent cue	.53	.02	.26	.18	.11

criterion of one errorless run in initial learning) and nonlearners. Overall, nonlearners averaged .77 more errors on Face 4 than Face 7 during learning, $t(59) = 1.72$, $p < .10$, and learners averaged 1.22 more errors on Face 4, $t(35) = 2.69$, $p < .05$. The larger difference for learners is to be expected, since differences should not begin to appear until at least a modest amount of learning has taken place. Finally, informal attempts to fit the transfer performance of learners and nonlearners separately suggest that differences can be fairly accurately described simply in terms of differences in the parameters of the context model (the similarity parameters for nonlearners are higher).

Although some learning is needed before differences between Faces 4 and 7 are detectable, the relative difficulty of the two faces did not interact with stage of practice in any way that suggests that the context model describes performance at one stage of learning and independent-cue models are correct for a different stage. Breaking the 32 original-learning trials into four blocks (of eight trials each) yielded an average number of errors of 2.06, 1.72, 1.08, and .84 for Face 4 and an average of 2.02, 1.24, .90 and .66 for Face 7. Thus Face 7 led to better performance than Face 4 on each block.

Discussion

The main results are easy to describe. The instructional variations produced large differences in the pattern of errors, reaction time, and transfer performance. There were strong interactions of instructional conditions with particular faces. Yet certain relationships in the data held across all conditions, relationships that were accurately described by the context model. Furthermore, this model provided a good account of many of the instructionally induced differences simply in terms of variations in the similarity parameters of the attributes. No new or special processes were required for the different conditions, and the context model fit the data better than an independent-cue model in each condition. At least for the present studies, instructional manipulations influence the representations but not the basic processes operating on them. Despite the variations in performance for each of the instructional conditions, performance was more in line with interactive-cue models than with models assuming that information is combined in an additive and independent manner. This raises the possibility that it may not be necessary for a new process model to be developed for each alternative strategy a person might employ. One might opt for formulating process models on a level at which they can capture the relations in the data that are invariant over strategy. There is little point in speculating about the viability of extending the context model to still other strategy variations, but we know from the present study that at least some degree of generality can be achieved even when experimental manipulations dramatically alter many details of performance.

As noted earlier one should also be cautious about generalizing the present results to different stimuli and different category structures, and at best the present results should be taken as suggestive. Still there is evidence that the advantage of interactive models over independent-cue models could have some generality. The Medin and Schaffer (1978) results held across both face and geometric stimuli and across at least modest variations in category structure. In none of these experiments, however, has the number of alternative training stimuli been very large. In a recent series of studies reported by Medin (Note 2), category size was varied in a major way. These experiments compared the difficulty of learning categories that either were or were not linearly separable. Independent-cue models predict that with other factors held constant, the linearly separable task should be easier. In different experiments the stimulus set was either small or unlimited (no stimulus was ever repeated). In neither case was there any evidence that the linearly separable task was easier, contrary to independent-cue models.

The present approach to the role of strategies in learning departs from usual practices. That is, typically attention is focused directly on strategies rather than on the byproducts associated with the use of strategies. And usually one is confronted with, and takes as the appropriate task, addressing the diversity and flexibility of strategies. By concentrating on the representations that result from the use of strategies, attention is called to the commonalities underlying performance. The present data are consistent with the idea that a basic property of categorization is that probes act as retrieval cues to access representations similar to the probe. This sensitivity to similarity acts as an efficient mechanism for categorization by analogy.

Brooks (1978) has argued that analogical reasoning has been given short shrift in analyses of categorization learning in favor of attention to the more analytical thinking associated with strategies and hypothesis testing. Indeed, there is some evidence that for complex category structures, mastery of the category is accomplished better in the absence of strategies than in their presence (Kossan, 1978; Reber, 1969, 1976). The context model suggests that analytical and analogical processes should not be viewed as mutually exclusive alternatives. Rather, access to stored representations may always be by means of an essentially analogical process, but the character of the representations may be modified by analytical strategies employed during learning.

Reference Notes

1. Shiffrin, R. M. Personal communication, January 26, 1981.
2. Medin, D. L. *Levels of categorization and the integration of new experiences.* Paper presented at the meeting of the American Psychological Association, Montreal, Canada, September 1980.

References

Brooks, L. Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, N.J.: Erlbaum, 1978.

Bourne, L. E., & O'Banion, K. Memory of individual events in concept identification. *Psychonomic Science,* 1969, *16,* 101–103.

Calfee, R. C. Recall and recognition memory in concept identification. *Journal of Experimental Psychology,* 1969, *81,* 436–440.

Franks, J. J., & Bransford, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology,* 1971, *90,* 65–74.

Hintzman, D. L., & Ludlam, G. Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition,* 1980, *8,* 378–382.

Kossan, N. E. *Structure and strategy in concept acquisition.* Unpublished doctoral dissertation, Stanford University, 1978.

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

Posner, M. I., & Keele, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology,* 1968, *77,* 353–363.

Reber, A. S. Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology,* 1969, *81,* 115–119.

Reber, A. S. Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, *2,* 88–94.

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology,* 1972, *3,* 382–407.

Received December 10, 1980 ■