Journal of Experimental Psychology:
Learning, Memory, and Cognition
1982, Vol. 8, No. 1, 37-50

0096-1515/82/0801-0037$00.75

Correlated Symptoms and Simulated Medical Classification

Douglas L. Medin
University of Illinois at Urbana-Champaign

Mark W. Altom
Rutgers—The State University

Stephen M. Edelson and Deborah Freko
University of Illinois at Urbana-Champaign

Category learning theories can be separated into those that expect judgments to be sensitive to configural information and those that expect judgments to be based on a weighted, additive summation of information. Predictions of these two classes of models were investigated in a simulated medical diagnosis task. Subjects learned about a fictitious disease or about two diseases from hypothetical case studies in which some symptoms were correlated with each other and others were independent. Following this initial training, subjects were presented either with pairs of new cases and asked to judge which was more likely to have the disease or with a single case and asked which disease was present. Across four experiments, subjects proved to be sensitive to configural information. When choosing between pairs of new cases, subjects tended to choose the case that preserved the correlation over the case that broke the correlation, even when the case with correlated symptoms contained fewer typical symptoms. When judging which disease was present in a single case, subjects' diagnoses were determined primarily by the correlated symptoms. Implications of these findings to process models of categorization are discussed.

Considerable evidence suggests that many natural categories are "fuzzy" and contain no simple set of defining features (for a recent review, see Mervis & Rosch, 1981). Consequently, attention has turned to how different sources of information (i.e., attributes) are integrated in order to make category judgments. When people learn about a category, do they learn about each attribute independently, or do they instead learn about relations that hold between attributes? The problem is of special interest under conditions in which category structure is not directly specified but must instead be based on experience with exemplars of a category.

In this study, classification into ill-defined categories was investigated in two types of simulated medical diagnosis tasks. In the single disease task used in Experiments 1–3, subjects learned about a fictitious disease from hypothetical case studies of patients having the disease. The case studies included descriptions of symptoms that tended to be characteristic of the disease. Some symptoms were correlated with each other, whereas others were independent. After subjects studied the descriptions, they were presented with pairs of new cases and asked to judge which was more likely to have the disease based on what they had learned from the earlier case studies. The second medical diagnosis task, used in Experiment 4, involved case studies from two contrasting hypothetical diseases, and again some symptoms were correlated and others were not. Following an initial learning phase, subjects were required to diagnose new cases. Across the four experiments, the principal question that was investigated concerned the extent to which subjects' diagnoses of new cases were sensitive to the correlations between symptoms.

In considering alternative categorization models that might be useful in investigating this question, theories can be separated into two classes—those that assume judgments are sensitive to correlated symptoms and those that assume the symptoms are treated

This research was supported in part by National Institute of Mental Health Grant MH32370 to Douglas L. Medin and by a biomedical research support grant from Rutgers University to Mark W. Altom.

Requests for reprints should be sent to Douglas L. Medin, Department of Psychology, University of Illinois, 603 E. Daniel, Champaign, Illinois 61820.

as independent. Theories of abstract prototype formation fall in the class of independent cue models (Medin & Schaffer, 1978). In these models, it is assumed that people represent the category in terms of a category prototype abstracted from the characteristic or central tendency of the attributes of the category members (cf. Posner & Keele, 1968). Category judgments of new items are based on their similarity to the abstracted category prototype. Since similarity to candidate prototypes has been assumed to arise from weighted, additive combinations of attributes composing the prototypes (Reed, 1972, 1973), classification judgments are predicted to be insensitive to correlations.

The context model of classification recently proposed by Medin (1975; Medin & Schaffer, 1978) provides a process model that assumes that people are sensitive to the configural information provided by correlated attributes when they make classification judgments. According to the context model, the item to be categorized acts as a retrieval cue that provides access to memory representations of one or more known exemplars. The greater the similarity of the unknown instance to an exemplar represented in memory, the greater the probability that the exemplar will be retrieved. Category judgments are based on the category membership of the retrieved exemplars. Sensitivity to correlated attributes is an indirect result of memory retrieval processes. Similarity values of the attributes comprising the item to those of the memory representations of the category exemplar are combined in a multiplicative fashion to determine overall similarity. This multiplicative assumption implies that cases that preserve a correlation between attributes should be more likely to retrieve exemplars represented in memory than cases that do not preserve the correlation.

The context model is just one example of a class of models referred to by Medin and Smith (1981) as relational coding models. These theories have in common the implication that component dimensions are not treated as additive and independent. We phrase our discussion in terms of the context model because it permits us to make detailed quantitative predictions, but it should be remembered that results consistent with the context model will be at least qualitatively consistent with all relational coding models.

Our primary concern is with the role of correlated symptoms in our simulated medical diagnosis task. Relational coding models predict that people will be sensitive to correlated symptoms; in contrast, independent cue models have no basis for predicting such sensitivity.

Experiment 1

The first experiment used two groups of subjects, one given training with cases that included a correlation between two symptoms and the other given training with cases in which the symptoms were all independent. During an initial training phase, subjects were presented with nine different case studies of the fictitious disease, burlosis. A given patient could have from one to four symptoms, and each symptom appeared in six of the nine patient descriptions.

The basic design of the learning cases for the correlated case study condition is depicted in Figure 1. An abstract notation of 1s and 0s is used to represent presence or absence, respectively, of each symptom. For example, the code 1100 would represent presence of the first two symptoms and absence of the last two symptoms. Two symptoms were perfectly correlated in that a given patient either had both symptoms or neither symptom. For example, in Figure 1, discolored gums and nosebleeds were perfectly correlated symptoms. Assignment of symptoms to the correlation condition was counterbalanced across subjects.

Following the learning phase, subjects received a series of pairs of cases. In each pair, subjects were instructed to choose the case that was more likely to have burlosis. In general, both independent cue and relational coding models predict that the case with the greater number of symptoms will be chosen as the one more likely to have burlosis. However, relational coding models (as represented by the context model) differ from independent cue models in their predictions on cases that do and do not preserve the correlation between symptoms. Suppose the test

SYMPTOMS OF BURLOSIS–CORRELATED CONDITION

CASE STUDY	SWOLLEN EYELIDS	SPLOTCHES ON EARS	DISCOLORED GUMS	NOSE-BLEED
1. R.C.	1	1	1	1
2. R.M.	1	1	1	1
3. J.J.	0	1	0	0
4. L.F.	1	1	1	1
5. A.M.	1	0	1	1
6. J.S.	1	1	0	0
7. S.T.	0	1	1	1
8. S.E.	1	0	0	0
9. E.M.	0	0	1	1

Figure 1. Learning cases represented in abstract notation for the correlated condition of Experiments 1 and 2.

pits a patient with a symptom pattern 1101 against a patient with a pattern 0111. The context model predicts that subjects should judge that the latter patient (where the correlation is preserved) is more likely to have burlosis than the former patient. Independent cue models could predict a difference if more weight were given to the third and fourth symptoms than to the first two symptoms. In practice, however, symptoms were counterbalanced, so that there is no a priori basis for expecting differential weighting. Even with differential weighting of the two correlated symptoms, other tests distinguish independent cue models from the context model. For example, differential weighting would lead one to expect that a patient with a symptom pattern 1010 is more likely to have burlosis than a patient with a symptom pattern 1100. The context model leads to the opposite prediction.

As an additional control, a second group of subjects was trained with the pattern of case studies represented in Figure 2. Although there are no correlations between the symptoms, the overall frequency of each symptom is the same as in Figure 1. Subjects in this uncorrelated case study condition received the same test cases as subjects in the correlated case study condition. Thus subjects trained on the cases in Figure 2 provide a baseline for evaluating effects of correlated symptoms. If subjects represent the disease in terms of the prototypical symptoms abstracted from the case studies, then diagnoses of the test cases should be similar for subjects in the two training conditions. In contrast, if subjects make their judgments in accord with the assumptions of the context model, then the patterns of diagnoses should be very different for the two training conditions, with only subjects trained with the cases in Figure 1 showing sensitivity to whether the test case preserves the correlation.

Method

Subjects. Subjects were 48 male and female undergraduate students enrolled in an introductory psychology course at the University of Illinois. Participation in the study partially fulfilled a course requirement. They were informed of their rights as participants and, at the end of the approximately 40-min. experimental session, were given an extensive oral and written description of the purpose of the experiment. Subjects were randomly assigned to either the correlated case study condition or the uncorrelated case study condition.

Stimuli. All cases consisted of patient descriptions typed on plain white index cards. These descriptions included the patient's name, age, occupation, and symptoms of the fictitious disease, burlosis. The nine learning cases indicated the presence or absence of each of the four symptoms according to the binary notation system shown in Figures 1 and 2. Fourteen of the transfer case descriptions indicated the presence or absence of all four symptoms (full information test cards), whereas the remaining 16 transfer cases indicated only the presence or absence of two of the symptoms and stated that no information was available concerning the remaining two symptoms (partial information test cards). The specific patterns of symptoms for the transfer cases are given in Table 1.

SYMPTOMS OF BURLOSIS-UNCORRELATED CONDITION

CASE STUDY	SWOLLEN EYELIDS	SPLOTCHES ON EARS	DISCOLORED GUMS	NOSE-BLEED
1. A.M.	1	1	0	1
2. S.E.	1	1	1	0
3. R.L.	1	1	1	1
4. L.F.	1	0	0	1
5. R.M.	1	1	1	1
6. J.S.	1	0	1	0
7. S.T.	0	1	1	1
8. J.J.	0	1	0	0
9. E.M.	0	0	1	1

Figure 2. Learning cases represented in abstract notation for the uncorrelated condition of Experiments 1 and 2.

Although all subjects in a group were given the same abstract pattern of descriptions and tests, the particular symptom assignments were randomized across subjects. For example, discolored gums was Symptom 1 for some subjects; for others it was Symptom 2, 3, or 4. Six different orders of symptoms were constructed such that each possible pair of symptoms served as the correlated symptoms for the correlated group and were Symptoms 3 and 4 for the uncorrelated group. Four subjects in each group received a given order of symptoms. Even when a symptom pattern was repeated for a patient description in the learning and transfer phases, a different name, age, and occupation was used, so that a given subject was presented with 39 different hypothetical patients.

Design. For subjects in the correlated group (Figure 1), Symptoms 3 and 4 were perfectly correlated during the training phase. No correlations between symptoms were present for the uncorrelated group. In both conditions, each symptom was present in six of the nine patients described during the learning phase. The symptoms were nosebleeds, swollen eyelids, discolored gums, and red splotches on ears.

The test descriptions were either full or partial descriptions. Six of the seven full information tests pitted a description having the same values on Symptoms 3 and 4 (preserving the correlated symptoms for the correlated group) against a description having different values for Symptoms 3 and 4. Number of symptoms was the same for each of the contrasts. The first four partial information tests pitted information on the first two symptoms against information on the last two symptoms. The other tests were included primarily to conceal the main purpose of the experiment.

Procedure. Subjects participated in the experiment in individual sessions. They were presented with the nine learning case descriptions and asked to study them for 20 min. in order to learn as much as possible about burlosis. Subjects were told that they could rearrange the cards in any fashion to best study them. After this initial learning phase, the cards were taken away from the subjects and the experimenter presented pairs of test cases, asking the subjects to indicate which of the two patients in each pair was more likely to have burlosis. The seven full information tests were presented first in random order, which varied across subjects. After explaining that the next series of tests would not have complete information, the experimenter presented the eight partial information tests, again in a random order, which varied across subjects. No feedback was given concerning the correctness or appropriateness of any of the subjects' judgments during the transfer test phase.

At the end of the experiment, subjects were asked to write what they thought a typical case of burlosis would be and to list the strategy they used in deciding who was more likely to have burlosis.

Results

Frequencies of choices of each alternative during the transfer phase are presented in the left half of Table 1. On the six critical

Table 1
Test Descriptions Used in Experiments 1 and 2, and Responses on the Tests by the Correlated and Uncorrelated Groups

Test pair		Experiment 1				Experiment 2			
		Correlated		Uncorrelated		Correlated		Uncorrelated	
a vs.	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Full information tests									
1101	0111	6	18	14	10	4	20	10	14
1110	0111	10	14	15	9	8	16	12	12
1101	1011	8	16	15	9	6	18	14	10
1110	1011	10	14	13	11	7	17	16	8
1010	1100	8	16	9	15	4	20	12	12
0001	0100	8	16	9	15	5	19	10	14
1100	0011	12	12	14	10	18	6	11	13
Partial information tests									
11--	--11	14	10	11	13	6	18	12	12
00--	--00	6	18	9	15	3	21	9	15
10--	--01	20	4	16	8	18	6	11	13
01--	--10	20	4	17	7	20	4	12	12
0-1-	1--1	6	18	8	16	4	20	6	18
1--0	10--	7	17	12	12	16	8	13	11
-1-0	0-1-	15	9	18	6	7	17	10	14
-0-1	1-0-	11	13	7	17	12	12	8	16

full information cases, subjects in the correlated group were much more likely than subjects in the uncorrelated group to choose the alternative with the same values on Symptoms 3 and 4 (*b* cases) than the alternative with different values on Symptoms 3 and 4 (*a* cases). That is, subjects in the correlated condition were sensitive to the correlated attributes in making their judgments. Overall, subjects in the correlated group chose the case that preserved the correlation 65% of the time compared with 48% of the time for subjects in the uncorrelated group. A *t* test combining responses over the six tests indicated that this difference was significant, $t(46) = 2.26$, $p < .05$.

The partial information tests were less revealing. For the two tests that pitted different values on Symptoms 1 and 2 against different values on Symptoms 3 and 4 (10-- vs. --01, and 01-- vs. --10), subjects in the correlated group chose the description with different values for Symptoms 1 and 2 83% of the time; subjects in the uncorrelated group showed a slightly smaller but similar preference (69%). An analysis of variance on the first four partial information tests using the factors of group and test type (correlation preserved vs. broken) was conducted. The Group × Test interaction was not significant, $F(1, 46) = 2.21$, and only the main effect of test type was reliable, $F(1, 46) = 49.02$, $MS_e = .231$, $p < .01$.

Nevertheless, this pattern of choices by the correlated group rules out the possibility that response preferences on the full description tests can be explained in terms of Symptoms 3 and 4 simply being weighted more heavily than Symptoms 1 and 2. For tests pitting the same values on Symptoms 3 and 4 against the same values on Symptoms 1 and 2, there were no overall group differences. Finally, on the one test in which the number of symptoms in the contrast varied (0-1- vs. 1--1), subjects in both groups selected the patient with more symptoms as more likely to have the disease [across groups, overall $\chi^2(1) = 8.33$, $p < .01$].

Analysis of subjects' descriptions of their strategies indicated that they often reported correlations that did not exist. Of the 24 subjects in the correlated group, 8 reported finding the correct correlation, but 4 other subjects reported symptom correlations that did not exist. In addition, 2 subjects reported using age–symptom correlations, 2 tried occupation alone, and 3 attempted to use occupation–symptom correlations in deciding who had burlosis, believing that these factors were related to the disease. In the uncorrelated group, 7 subjects reported finding correlated symptoms, 2 reported using age alone, 2 attempted to use age–symptom correlations, 1 used an age–occupation correlation, 4 used occupation alone, and 3 reported occupation–symptom correlations.

Discussion

The results were clear-cut. Subjects in the correlated group did not behave in accord with independent cue models in that they consistently indicated that patient descriptions preserving the correlation between symptoms were more likely to have the disease than patient descriptions that did not preserve the correlation. Subjects in the uncorrelated group did not show this preference. The results cannot be explained by the suggestion that the correlated symptoms independently became salient, because the partial information test descriptions with different values (one present, one absent) on Symptoms 3 and 4, which had been correlated, were judged as less likely to have burlosis than descriptions with different values on Symptoms 1 and 2.

The results are compatible with the context model's prediction that people should be sensitive to correlated attributes or symptoms. In fact, a considerable percentage of subjects reported noting correlations between symptoms that were not actually present. Use of nonexistent or illusory correlations has been noted before in the literature on diagnosis (Chapman & Chapman, 1967, 1969), and it suggests that people are biased toward perceiving configural relations, even when such configurations are absent.

Experiment 2

In the first experiment, a given symptom was either present or absent. This practice produced some rather artificial transfer tests. For example, a subject might have been asked, "Who is more likely to have burlosis, a person with normal gums and no nosebleeds or a person without splotches on their ears and normal eyelids?" The second experiment employed the case study structure of the first experiment, but instead of a symptom being present or absent, symptoms were organized along four binary dimensions. That is, a patient either had an elevated white blood cell count or an elevated red blood cell count, splotches on the skin or a skin rash, puffy eyes or sunken eyes, and high blood pressure or low blood pressure. These symptoms were arranged in accordance with Figures 1 and 2, but now notation 1 referred to the more frequent or typical symptom and 0 referred to the less frequent or atypical symptom. For example, six of the nine case studies during the learning phase might have puffy eyes as a symptom and the other three would have sunken eyes. As in Experiment 1, both correlated and uncorrelated conditions were used. For the correlated condition, the two correlated symptoms for a given case study had either the two typical complaints or the two atypical complaints.

The main purpose of the second experiment was to replicate the findings of the first experiment. The procedure was modified to produce more plausible transfer tests. The task in the second experiment corresponded to a situation in which patients were clearly ill and the main question was whether they had the particular disease under consideration.

Method

Subjects. The subjects were 48 male and female undergraduate students enrolled in an introductory psychology course at the University of Illinois, who participated in partial fulfillment of a course requirement. Subjects were informed of their rights as participants and, at the end of the aproximately 40-min. experimental session, were given an extensive oral and written description of the purpose of the experiment. Participants were randomly assigned to either the correlated case study condition or the uncorrelated case study condition.

Stimuli. The case study descriptions were directly analogous to those used in the first experiment, except that all descriptions included four symptoms that could either be typical (present in six of the nine cases) or atypical (present in three of the nine cases). That is, atypical symptoms took the place of the absence of a symptom. The symptoms were elevated white blood cell count or elevated red blood cell count, splotches on the skin or skin rash, puffy eyes or sunken eyes, and high blood pressure or low blood pressure. Blood cell count, skin condition, eye condition, and blood pressure were considered as four symptom dimensions; one form of complaint (e.g., skin rash) was typical and frequent (designated by a 1 in Figures 1 and 2), and the other form (e.g., splotches on skin) was atypical and infrequent (designated by a 0 in Figures 1 and 2).

As in the first experiment, there were six orders of symptom dimensions to counterbalance the correlated symptoms. In addition, the two complaints on each symptom dimension were alternated so that each complaint was typical in three of the six orders. To control for symptoms always being typical or atypical together, only two dimensions were alternated at a time. For example, the pattern 1111 for some subjects meant "high blood pressure, red blood cell count up, puffy eyes, and skin rash," whereas for others 1111 meant "low blood pressure, white blood cell count up, puffy eyes, and skin rash."

Design and procedure. All details of design and procedure exactly correspond to those of the first experiment. Subjects studied case studies conforming either to the design in Figure 1 or the design in Figure 2, and then were given transfer comparisons corresponding to those in Table 1. The only difference is that the notation 0 refers not to the absence of a symptom, but rather to the presence of the atypical symptom.

Results

Frequencies of choices on the transfer tests are given in the right half of Table 1. Subjects in the correlated group were again much more likely than subjects in the uncorrelated group to judge that a test description that preserved the correlated symptom pattern was more likely to have burlosis than a test description that did not preserve the correlation. This was true even though the pairs of descriptions had the same number of characteristic or typical symptoms. On the full information tests, subjects in the correlated group chose the description preserving the correlation 76% of the time compared to 49% for the uncorrelated group. This difference is highly reliable, $t(46) = 3.69$, $p < .01$.

The same pattern held for the partial information tests. The first two partial information comparisons show that typical correlated symptoms were judged more likely to be associated with burlosis than two typical uncorrelated symptoms, and that two atypical correlated symptoms were considered to be more diagnostic than two atypical uncorrelated symptoms for subjects in the correlated group. The third and fourth tests revealed that a description involving one typical and one atypical symptom that broke the correlation was judged to be less likely to indicate burlosis than an atypical and a typical symptom that had not been correlated for subjects in the correlated group. A 2×2 analysis of variance was conducted on these tests using the factors of group and test type (correlation broken vs. preserved). Both the main effect of test type, $F(1, 46) = 19.29$, $MS_e = .519$, $p < .01$, and the theoretically important interaction of group with test type, $F(1, 46) = 14.63$, $MS_e = .519$, $p < .01$, were highly significant. Finally, on the test pitting one typical and one atypical symptom against two typical symptoms, subjects in both groups chose the description with two typical symptoms as being more likely to indicate burlosis, χ^2 (1) = 16.3, $p < .01$.

Subjects' written descriptions again revealed reports of nonexistent correlations. Of the 24 participants in the correlated condition, 14 reported noticing the correct correlation. In the uncorrelated group, 4 subjects noted nonexistent symptom correlations, 3 reported occupation–symptom correlations, 3 noted age–symptom correlations, 1 noted an age–occupation correlation, and 4 subjects reported occupation alone in deciding who had burlosis.

Discussion

The results replicate those of the first experiment. Subjects in the correlated condition both noticed and used correlated symptoms. Hypothetical patients whose symptom pattern preserved the symptom correlations were judged more likely to have burlosis than hypothetical patients whose symptom pattern broke the symptom correlation. This use of correlated symptoms is consistent with relational coding models, such as the context model, and inconsistent with the class of independent cue models. The next experiment further explores this use of correlated symptoms under conditions in which test cases involve a trade-off between symptom correlations and number of typical symptoms.

Experiment 3

In virtually all of the contrasts of interest in the first two experiments, either the number of symptoms or the number of characteristic symptoms was held constant. In Experiment 3, transfer comparisons were made involving cases with different numbers of characteristic symptoms. Independent cue models imply that people will select descriptions with a greater number of characteristic symptoms as more likely to manifest burlosis. Predictions of relational models depend on whether patterns of correlated symptoms are preserved.

The design of the case studies used in the learning phase is shown in Figure 3. There are five binary valued symptom dimensions, and for each dimension the typical symptom appears in six of the nine cases. The symptom pattern for the fourth and fifth symptom dimension is perfectly correlated in that a case study either has typical symptoms on both dimensions or atypical symptoms on both dimensions.

The transfer comparison tests are shown in Table 2. The first nine comparisons are of primary interest, because in each test, the option labeled *b* has more characteristic symptoms than the option labeled *a*. Consequently, independent cue models predict

SYMPTOMS OF BURLOSIS

CASE STUDY	BLOOD PRESSURE	SKIN CONDITION	MUSCLE CONDITION	CONDITION OF EYES	WEIGHT CONDITION
1. R.L.	0	1	0	1	1
2. L.F.	1	1	0	1	1
3. J.J.	0	0	1	1	1
4. R.M.	1	0	1	1	1
5. A.M.	1	1	1	1	1
6. J.S.	1	1	1	1	1
7. S.T.	1	0	0	0	0
8. S.E.	0	1	1	0	0
9. E.M.	1	1	1	0	0

Figure 3. Learning cases represented in abstract notation for Experiment 3.

Table 2
Transfer Tests and Response Proportions From Experiment 3

Test Pair *a*	vs. *b*	Proportions of choices of *a* alternative: Observed	Predicted
01110	11101	.37	.40
11001	11110	.43	.47
01010	11010	.50	.47
10001	10101	.23	.32
11100	11101	.57	.56
00111	11101	.57	.65
01011	11110	.70	.66
00100	00101	.53	.48
10000	10010	.63	.66
11000	01100	.33	.30
10100	01100	.33	.40
10011	11100	.50	.42
10011	01011	.40	.40
01111	11011	.60	.58
00011	01100	.37	.41
00100	10000	.57	.53
11100	01011	.50	.49
00111	01011	.60	.59
10111	11011	.53	.60
11001	01101	.40	.41

that the *b* case description will be judged to be more indicative of burlosis than the *a* case description. Relational coding models do not share this prediction, at least for comparisons five through nine. On these tests, the description with fewer characteristic symptoms preserves the correlation between the fourth and fifth symptom dimensions, whereas the alternative description breaks the correlation. Therefore, these models predict that the *a* option—that is, the option with fewer characteristic symptoms—generally will be selected on these tests. The remaining tests in Table 2 involve cases with equal numbers of typical symptoms and were included to disguise the major purpose of the experiment and to supply additional observations for detailed parameter estimates.

Exact predictions of the context model depend on the salience of the different symptom dimensions, and to fully evaluate the predictions of the context model, we attempted quantitative fits to the data. According to the context model, when two items are presented on a transfer test they will act as retrieval cues to access stored information associated with similar case studies. Similarity along a symptom dimension is assumed to vary between 0 and 1, with 1 representing identity or maximum similarity. Overall similarity of a transfer item is assumed to be a multiplicative function of individual symptom similarities. If we represent the similarity on the five symptom dimensions in Figure 3 by *a*, *b*, *c*, *d*, and *e*, respectively, then transfer description 01110 would have an overall similarity to case study R. L. (01011) of $1 \cdot 1 \cdot c \cdot 1 \cdot e$, or $c \cdot e$. By the same system, the new item would have similarity $a \cdot c \cdot e$ to case study L. F. (11011), $b \cdot e$ to case study J. J. (00111), and so on. The smaller the value of similarity between the typical and atypical symptoms, the more salient the symptom dimension. To generate transfer test predictions, the five similarity parameters are estimated from the data, and the probability of selecting the *a* choice as more likely to have burlosis is taken to be simply the sum of the similarities of the *a* choice to each of the case studies used in training divided by that quantity plus the sum of the similarities of the *b* choice to each of the nine case studies.

Method

Subjects. The subjects were 30 male and female students attending the University of Illinois. They were paid $3 for the approximately 40-min. experimental session. Subjects were informed of their rights as participants and, at the end of the session, were given an extensive oral and written description of the purpose of the experiment.

Stimuli. The case study descriptions were prepared in a manner analogous to that of the first two experiments. Each description involved five symptom dimensions: blood pressure (high or low), skin condition (splotches or rash), muscle condition (stiffness or loss of control), eye condition (puffy or sunken), and weight condition (loss or gain). One symptom of each pair was selected to be typical (1 in Figure 3), and the other was selected to be atypical (0 in Figure 3). Names, ages, and occupations were unique to each description, and learning and test descriptions conformed to the structure indicated in Figure 3 and Table 2, respectively.

Design and procedure. Details of design and procedure were similar to those of the first two experiments. However, in this experiment only a single group was used, and for that group Symptom Dimensions 4 and 5 were perfectly correlated, as indicated in Figure 3. Symptom order was randomized across subjects, and typical symptoms were alternated. Specifically, 10 different orders were prepared such that each possible pair of symptom dimensions was assigned to Symptom Dimensions 4 and 5 (the correlated dimensions) exactly once. After subjects studied the case descriptions in Figure 3, they were tested on the transfer descriptions indicated in Table 2.

Results

Predicted and observed proportions of choices of the *a* alternative on each transfer test are given in Table 2. In the first four tests, the pattern of symptom correlations was broken in both descriptions, and the description having the greater number of typical symptoms was judged more likely to be associated with burlosis. In the next five tests, in which the descriptions with fewer typical symptoms preserved the pattern of symptom correlations and the description with more typical symptoms did not, the pattern of choices was reversed. This interaction is predicted by relational coding but not by independent cue models. A *t* test on the proportion of *a* responses on Tests 1–4 versus tests 5–9 indicated that this difference was statistically reliable, $t(29) = 3.87$, $p < .01$.

As can be seen in Table 2, quantitative fits of the context model captured the major trends in the observed choice proportions. Best fitting similarity parameters were found using a least squares criterion. The resulting estimates for *a*, *b*, *c*, *d*, and *e* were .81, .61, .09, .44, and .08, respectively. Although each data point is based on only 30 observations, the average absolute deviation of predicted and observed proportions is .04, the root mean square deviation is .05, and 81% of the variance is accounted for by the model. One can readily see that the model predicts that the *b* option will be preferred on Tests 1–4, that the *a* option will be preferred on Tests 5–9, and that the magnitude of the predicted values closely follows the observed values.

Analysis of the written protocols revealed that 11 subjects reported the true symptom correlation. There were also six descriptions involving spurious symptom correlations, four reporting age–symptom correlations, and two reporting occupation–symptom correlations. Two subjects believed occupation was the determining factor of burlosis, and one subject felt that age was the determinant.

Discussion

For the third time in three experiments, subjects noticed and used correlated symptoms in their diagnoses. Even when it was the alternative with fewer typical symptoms, subjects tended to choose the case that preserved the correlation between symptoms over the alternative that broke that correlation. This finding is strong evidence against independent cue models, which predict that number of typical symptoms should be the principal determinant of performance. However, the fact that correlated symptoms had a large influence on subjects' judgments is readily accounted for within the context model, which fits the data both qualitatively and quantitatively.

Experiment 4

We have argued that the pattern of results across the first three experiments suggests that subjects are sensitive to correlated attributes when they make classification judgments. However, our use of a single category during category acquisition restricts the generality of our findings. For a number of natural categories (e.g., language acquisition), no obvious contrasting categories are available. Nevertheless, considerable evidence suggests that when contrast categories are available, they play a significant role in classification judgments (Rosch & Mervis, 1975; Tversky & Gati, 1978). It may very well be the case that mechanisms of classification consistent with independent cue models only come into play when contrasting categories are present to allow the learning of cue validities.

A second design feature of the first three experiments concerns the possibility that recognition of study cases influenced subjects' choices rather than sensitivity to correlations. Although all transfer items involved new names, ages, and occupations, almost without exception each transfer description that preserved the correlation between symptoms had an overall symptom

TERRIGITIS

CASE	D_1	D_2	D_3	D_4
E.M.	1	1	1	1
S.T.	0	1	1	1
R.M.	1	1	0	0
L.F.	1	0	0	0

MIDOSIS

CASE	D_1	D_2	D_3	D_4
R.L.	1	0	1	0
A.M.	0	0	1	0
S.E.	0	1	0	1
J.J.	0	0	0	1

Figure 4. Learning cases represented in abstract notation for Experiment 4.

pattern that exactly matched the symptom pattern of one of the original case studies. Consequently, it is not clear whether subjects were using the correlations per se or simply matching entire symptom patterns (and matching correlated symptoms coincidentally).

The fourth experiment was designed specifically to extend our paradigm to a multiple category situation and to assess the effects of new cases that preserved correlated symptoms but did not match exactly any case study from training. The case studies, shown in Figure 4, involved one of two diseases—terrigitis or midosis. For the first two symptom dimensions, the value 1 was typical of terrigitis and the value 0 typical of midosis. For the third and fourth symptom dimensions, however, neither value was more typical of terrigitis or midosis. Instead, values along the third and fourth dimensions were perfectly correlated, with values --11 and --00 occurring only for terrigitis and values --10 and --01 occurring only for midosis. Following initial training, subjects were presented with new case studies and asked to classify them as having terrigitis or midosis. All 16 possible combinations of symptom patterns were presented on these transfer tests. Of these new patterns, 8 exactly matched one of the original case studies, and 8 did not.

If subjects use only the correlated symptoms, then all descriptions having values 11 or 00 on the third and fourth dimensions should be diagnosed as having terrigitis, regardless of whether the symptom pattern exactly matches a case study. If exact matches are crucial, and subjects use typical symptoms as a basis for classification in the absence of exact matches, then the values on the first and second symptom dimensions will control responses when the transfer pattern does not exactly match a case study from training. Finally, the context model falls somewhere in between. It predicts that classification is based on similarity to case studies, but does not assume that only exact matches influence performance. The multiplicative similarity function also implies a sensitivity to correlated symptoms for both exact matches and nonexact matches that preserve symptom correlations. The success of the context model's predictions can be assessed by how well it fits the transfer data quantitatively.

Method

Subjects. Subjects were 32 male and female undergraduate students enrolled in an introductory psychology course at the University of Illinois, who participated in partial fulfillment of a course requirement. They were informed of their rights as participants and, at the end of the approximately 40-min. experimental session, were given an extensive oral and written description of the purpose of the experiment.

Stimuli and design. All cases consisted of patient descriptions typed on plain white index cards, which included the patient's name, age, occupation, and pattern of symptoms. The symptom dimensions were weight (gain or loss), eyes (puffy or sunken), muscle state (stiffness or spasms), and skin condition (splotches or rash).

During the transfer tests, 32 new descriptions were presented. Each of the 16 possible combinations of the four binary-valued symptom dimensions was presented twice. All transfer cards were similar to the ones used in training, except that names, ages, and occupations were unique to each card.

Design of the training stimuli followed the abstract notation shown in Figure 4. Other details of counterbalancing, such as which particular symptoms were correlated, followed the procedures used in the first three experiments.

Procedure. Subjects were individually tested in a procedure that closely followed that of earlier experiments. They were first presented written instructions telling them to imagine themselves as doctors identifying two diseases, terrigitis and midosis. Subjects were also told to "develop an impression of both diseases and the symptoms involved." Subjects were instructed that the names, ages, and occupations listed on the cards were used to identify the patients and were not related to the diseases.

After reading the instructions, subjects were presented the eight learning cards, with a title card of each disease next to the four cards corresponding to the disease. Subjects studied the cards for 10 min. They were then given a learning test in which the experimenter mixed up all eight learning cards and told the subject to sort the cards into the appropriate disease category. Those subjects who could not successfully sort all the cards were given an additional 5 min. to study them, followed by a second learning test. The majority of subjects passed the learning test on their first try, and the remaining subjects passed on their second try.

Transfer tests immediately followed the learning phase. Subjects were presented with the 32 transfer cases, representing two descriptions of each of the 16 possible combinations of symptoms. Subjects were presented the cards in a random order, one at a time, and were told to indicate which disease they thought the patient had, based on what they had learned.

Results

The proportions of diagnoses of each transfer case as terrigitis are given in Table 3, broken down according to whether or not the new cases matched patterns of symptoms from the training phase. Each data point is based on 64 observations. Subjects were quite accurate on symptom patterns that exactly matched a case study from the training phase. There was also a systematic effect of the first two symptom dimensions; subjects were 12 to 16 percentage points more accurate if the first and second symptoms were typical of the disease than if one of the two symptoms was atypical.

Table 3
Transfer Tests and Response Proportions From Experiment 4

Test description	Proportions of terrigitis responses: Observed	Predicted
Exact match to old terrigitis patterns		
1111	.88	.85
1100	.89	.86
0111	.73	.76
1000	.77	.73
Exact match to old midosis patterns		
0010	.12	.15
0001	.17	.14
1010	.25	.24
0101	.33	.27
New patterns		
0000	.53	.56
0011	.53	.54
0100	.75	.73
1011	.67	.66
1110	.45	.46
1101	.38	.44
0110	.36	.34
1001	.28	.27

In general, new patterns were placed in the category associated with the symptom correlation that the new pattern manifested. That is, new patterns with values 11 and 00 on Symptoms 3 and 4 tended to be diagnosed as terrigitis, whereas new patterns with values 10 and 01 on Symptoms 3 and 4 tended to be diagnosed as midosis. In addition, there were clear effects of the first two symptom values. Thus, it appears that for both exact matches to old case studies and for new combinations of symptoms subjects used correlated symptoms as well as typical symptoms in making their diagnoses. This is precisely the pattern of performance predicted by the context model.

In order to apply the context model to this two-category situation, the probability of classifying a transfer description as terrigitis was assumed to equal the sum of similarities of the description to the four terrigitis case studies divided by that sum plus the sum of the similarities of the description to the four midosis case studies. Best fitting similarity parameters of the model were estimated according to a least squares criterion. The resulting estimates for a, b, c, and d were .52, .40, .32, and .06, respectively, and the predictions associated with these parameters are shown in Table 3. The parameter estimates suggest that the correlated dimensions represented by parameters c and d tended to be more salient than the first two dimensions. The fit of the model to the data is excellent—the average absolute deviation of predicted from observed proportions is less than .03, the root mean square error is .03, and 98% of the variance is accounted for by the model.

Discussion

Once again, subjects were sensitive to correlated symptoms, even when the transfer description did not exactly match a case study from original learning. In addition, this use of correlated symptoms did not preclude use of typical symptoms associated with the first two symptom dimensions. This pattern of performance is qualitatively consistent with relational coding models in general and is quantitatively consistent with the context model in particular.

General Discussion

Across the four experiments, subjects' judgments in the transfer phase were biased toward the alternatives that preserved the correlation between symptoms from the learning phase. In Experiment 1, subjects preferred test cases in which symptoms or lack of symptoms maintained a correlation from the learning phase. Subjects' choices on the partial information cases indicated that this bias could not be accounted for simply on the basis of the correlated symptoms independently receiving greater weight than the uncorrelated symptoms. In Experiment 2, the experimental design was extended to the situation in which the correlation between symptom dimensions was represented by the joint occurrence of two typical symptoms or two atypical symptoms. Again, subjects tended to choose test cases that maintained the correlation over test cases that did not.

Experiment 3 provided the most diagnostic evidence toward distinguishing between linear and configural models. When tests pitted cases maintaining the correlation against cases having more typical symptoms, subjects tended to choose the case that maintained the correlation. Finally, Experiment 4 extended the finding that subjects are sensitive to correlations between attributes to multiple category designs, in which contrasts between categories might be expected to influence acquisition. Moreover, Experiment 4 demonstrated that use of correlated symptoms was not an artifact of subjects relying on exact matches to patterns of symptoms associated with training case studies. In addition, subjects in these experiments often reported that they used nonexistent correlations. Use of spurious correlations as well as use of existing correlations suggests that the underlying decision process involves more than the weighted sum of independent attributes.

The first three experiments used just a single category, whereas most categorization studies involve at least two contrasting categories. It is therefore of some interest that subjects found the category formation task to be meaningful without the presence of a contrast category. The fourth experiment indicates that the use of correlated symptoms cannot be due to use of a single category. Instead, the results are consistent with an emerging body of evidence against independent cue models in categorization tasks (cf. Medin & Schaffer, 1978; Medin & Smith, 1981). The series of studies on linear separability by Medin and Schwanenflugel (1981) seem most pertinent to the present experiments. Since independent cue models assume that categorization is based on a summing of evidence from independent dimensions against a criterion, they formally imply that categories must be linearly separable or separable by a linear discriminant function (Sebestyen, 1962). *Linearly separable categories* is the formal term for categories in which category instances can be correctly identified on the basis of a weighted, additive combination of component information.

In contrast to independent cue models, relational coding models imply that configural information, represented by the similarity of exemplars to each other, should be the main factor controlling the difficulty of categorization. To test this prediction, Medin and Schwanenlugel (1981) compared two categorization tasks in which average between-category similarity was held constant. In one task the categories were linearly separable, and in the other task the categories were not linearly separable, but instances within a category were of high similarity. Subjects found the linearly separable task to be more difficult than the task with categories that were not linearly separable. Even when instances of high similarity were held constant, linear separability still did not emerge as a factor in classification learning.

Given that subjects are sensitive to correlations between attribute dimensions, several different underlying processes could account for subjects' judgments. The context model is just one of the class of relational coding models that imply that people should be sensitive to configural information. For example, several models of category formation (Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974, 1977; Reitman & Bower, 1973) assume that subjects accumulate counts of the numbers of times different subsets of features occur together. Judgments of category membership are then based on a summation of frequency scores. If subjects encode relational frequencies, as opposed to the frequencies of individual features, then sensitivity to correlated features would be a natural prediction. Correlated symptom dimensions would tend to have high counts for occurrence of both typical and both atypical symptoms, but low counts for occurrence of one typical and one atypical symptom.

These relational frequency models differ from the context model in their assumption that correlational information is directly represented in memory in the form of co-occurrence frequency counts. In contrast, the context model assumes that sensitivity to correlated features is an indirect result of memory retrieval processes. Test cases are assumed to act as cues that retrieve information from memory. The multiplicative similarity assumption implies that stimuli preserving a correlation between dimensions should be more likely to retrieve a case represented in memory than stimuli that do not preserve the correlation. The excellent quantitative fit to the data from the third and fourth experiments provides additional support for the context model. Unfortunately, direct comparisons with relational frequency models are precluded because relational frequency models have not yet been developed in sufficient detail to generate quantitative predictions.

The present series of studies provides a point of contact between work on clinical decision making (e.g., Dawes & Corrigan, 1974) and research on category formation. Independent cue models are formally analogous to linear decision models and can be considered as possible process models for clinical decision making (cf. Cantor, Smith, French, & Mezzich, 1980). Relational coding models parallel the use of configural information in clinical decision making (Einhorn, 1970). By providing candidate sets of contrasting process models, research on category formation may help sharpen the distinction between linear and configural models, as well as provide particularly diagnostic evidence for distinguishing these two classes of models for clinical decision making.

The present series of studies suggests that configural information does indeed play an important role in subjects' category judgments. Under conditions in which dimensions are perfectly correlated, subjects appear to encode and to use this configural information. Relational coding models imply that whether subjects are sensitive to configural information depends on the conditions of original learning. If the training cases include instances with correlations between the dimensions, then later judgments should be biased toward use of this configural information. In contrast, if the original training cases do not include correlations between the dimensions, then configural information may appear to play a negligible role in category judgments.

References

Cantor, N., Smith, E. E., French, R. de S., & Mezzich, J. Psychiatric diagnosis as prototype categorization. *Journal of Abnormal Psychology,* 1980, *89,* 181–193.

Chapman, L. J., & Chapman, J. P. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology,* 1967, *72,* 193–204.

Chapman, L. J., & Chapman, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology,* 1969, *74,* 271–280.

Dawes, R. M., & Corrigan, B. Linear models in decision making. *Psychological Bulletin,* 1974, *81,* 95–106.

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin,* 1970, *73,* 221–230.

Hayes-Roth, B., & Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior,* 1977, *16,* 321–338.

Medin, D. L. A theory of context in discrimination learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9). New York: Academic Press, 1975.

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

Medin, D. L., & Schwanenflugel, P. J. Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory,* 1981, *7,* 355–368.

Medin, D. L., & Smith, E. E. Strategies and classification learning. *Journal of Experimental Psychol-*

ogy: Human Learning and Memory, 1981, *7,* 241–253.

Mervis, C. B., & Rosch, E. Categorization of natural objects. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology,* 1981, *32,* 89–115.

Neumann, P. G. An attribute frequency model for the abstraction of prototypes. *Memory & Cognition,* 1974, *2,* 241–248.

Neumann, P. G. Visual prototype information with discontinuous representation of dimensions of variability. *Memory & Cognition,* 1977, *5,* 187–197.

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

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

Reed, S. K. *Psychological processes in pattern recognition.* New York; Academic Press, 1973.

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology,* 1973, *4,* 194–206.

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

Sebestyen, G. S. *Decision-making processes in pattern recognition.* New York: Macmillan, 1962.

Tversky, A., & Gati, I. Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, N.J.: Erlbaum, 1978.

Received April 4, 1981 ■

Instructions to Authors

The journal will favor long articles integrating the results of several experiments and/or aimed at important theoretical issues. Shorter articles will also be considered, however, when the results are particularly clear-cut and important to the field. A section entitled "Notes, Comments, and New Findings," is designed to contain contributions that do not meet the above criteria but that advance the field of experimental psychology in the areas of concern of this journal. Such contributions could include letters, comments on previously published articles or books, problems with theoretical conclusions or empirical designs of earlier articles, replies to criticism, and brief reports of new work that have particular merit but are not ready for publication as regular articles. For further information on content, authors should refer to the editorial in the July 1980 issue of this journal (Vol. 6, No. 4, pp. 439–440).

APA policy prohibits an author from simultaneously submitting the same manuscript to two or more journals. Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (2nd ed.). Instructions on tables, figures, references, metrics, and typing (all copy must be double-spaced) appear in the Manual. Manuscripts should include an abstract of 100–175 words. Authors are requested to refer to the "Guidelines for Nonsexist Language in APA Journals" (Publication Manual Change Sheet 2, *American Psychologist,* June 1977, pp. 487–494) before submitting manuscripts to this journal.

Authors should submit manuscripts in quadruplicate, and all copies should be clear, readable, and on paper of good quality. Authors should keep a copy of their manuscript to guard against loss. Mail manuscripts to the Editor, Richard M. Shiffrin, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

Addresses for the editors of the other JEP journals are as follows: *Journal of Experimental Psychology: General,* Gregory A. Kimble, Department of Psychology, Duke University, Durham, North Carolina 27706 (submit 4 copies of the manuscript); *Journal of Experimental Psychology: Human Perception and Performance,* William Epstein, Editor-elect, Department of Psychology, University of Wisconsin, W. J. Brogden Psychology Building, 1202 West Johnson Street, Madison, Wisconsin 53706 (submit 4 copies of the manuscript); *Journal of Experimental Psychology: Animal Behavior Processes,* Donald S. Blough, Department of Psychology, Brown University, Providence, Rhode Island 02912 (submit 3 copies of the manuscript). When one of the editors believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, he may redirect the manuscript with the approval of the author.

The editors have agreed to use blind review when it is requested by the author. Authors requesting blind review should prepare manuscripts to conceal their identity.