COGNITIVE PSYCHOLOGY 5, 1–28 (1973)

A Model for Analogical Reasoning[1]

DAVID E. RUMELHART AND ADELE A. ABRAHAMSON
University of California, San Diego

A theory of analogical reasoning is proposed in which the elements of a set of concepts, e.g., animals, are represented as points in a multidimensional Euclidean space. Four elements A,B,C,D, are in an analogical relationship A:B::C:D if the vector distance from A to B is the same as that from C to D. Given three elements A,B,C, an ideal solution point I for A:B::C:? exists. In a problem A:B::C:D_1, . . ., D_i, . . ., D_n, the probability of choosing D_i as the best solution is a monotonic decreasing function of the absolute distance of D_i from I. A stronger decision rule incorporating a negative exponential function in Luce's choice rule is also proposed. Both the strong and weak versions of the theory were supported in two experiments where Ss rank-ordered the alternatives in problems A:B::C:D_1,D_2,$D_3$$D_4$. In a third experiment the theory was applied and further tested in teaching new concepts by analogy.

Despite psychologists' considerable confidence that analogical reasoning plays an important role in intelligent behavior (cf., Bartlett, 1958; Miller, 1960; Minsky, 1966; Oppenheimer, 1956; Polya, 1957; and Reitman, 1964), analogy formation has received little systematic attention. Nearly all of the work done has been in the form of complex computer programs (cf. Becker, 1969; Evans, 1964; Reitman, 1966) without systematic comparisons of the programs with subject's behavior (see Hunt, 1968). The present paper outlines a simple theoretical model for the understanding of analogical reasoning and evaluates the validity of the model by means of empirical test.

To introduce our notion of analogical reasoning, it is useful to outline a definition of the word *reasoning* from which we can work. The term is used here to denote those processes in information retrieval which depend on the *structure*, as opposed to the *content* of organized memory.

[1] The research reported in this paper was supported in part by GB-8587 from the National Science Foundation, by Grant NG-07454-03 from the National Institute of Health, the United States Public Health Service. The paper was prepared for a Research Workshop on Cognitive Organization and Psychological Processes sponsored by the Committee on Basic Research in Education, National Research Council.
David E. Rumelhart, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

Thus, one might answer the question "Who is the father of your country?" in at least two different ways. In one case, the specific information that George Washington was the "father of our country" might be stored and used to answer the question. On the other hand, when specific information is not available, one can consult the stored meanings of the words in question and one's knowledge of history to derive a plausible answer. The first of these methods might be called remembering, since retrieval depends on the specific information stored. The second method may be identified with reasoning, since in this case retrieval depends to a much greater extent on the *form* of the relationship among the words. The same act of reasoning (i.e., the same processes) could have been applied to the question "Who was the father of your state?" or "Who was the mother of your country?". It is not the specific content of the question but the form of the relationships among the words which determines the response.

If one accepts this working definition of reasoning, then the theoretical problem in understanding any particular reasoning process becomes clear. We must (1) specify the form of the memory structure and then (2) determine the algorithm which is applied in the case of the reasoning process in question.

Perhaps the simplest reasoning task by our definition involves the judgment of similarity or dissimilarity of concepts. It is normally assumed that degree of similarity is not directly stored and thus "remembered" as such, but rather that it is derived from the memory structure. The simplest view holds that judged similarity between concepts is a simple function of the "psychological distance" between these concepts in the memory structure. The "closer" two concepts are to one another in memory, the more similar they are. Thus, in the case of judging similarities, the two questions which we need to answer are: (1) what is the nature of the memory structure which underlies similarity judgments, and (2) what is the measure of "distance" on this psychological space?

A recent paper by Nancy Henley (1969) illustrates one set of answers which have been given to these questions. Henley assumed (1) that the memory structure may be represented as a multidimensional Euclidean space and (2) that judged similarity is inversely related to distance in this multidimensional space. To test these assumptions in the semantic domain, Henley used the techniques developed by Shepard (1962), Kruskal (1967), and others to deduce the form of the space from subjects' judgments of similarity and dissimilarity among the concepts.

In one of her experiments, Henley used ratings of dissimilarity to deduce the underlying "psychological space" relating common animal terms to one another. She chose 30 of the most common mammals and asked subjects to rate all possible pairs of the 30 mammals as to their dissimilarity on a scale from 0 to 10. A value of 0 indicated that the animals were identical, a value of 10 indicated that they were maximally different. These data were used as input to TORSCA (a multidimensional scaling program developed by Young and Torgerson; 1967) to find the multidimensional solution with minimum dimensionality consistent with the observed dissimilarity data.

Henley's results showed that these 30 mammals fit reasonably well into a three-dimensional space (Kruskal's stress index at 9.4%). Figure 1 shows the three dimensions along with placements of some of the animals. Henley employed several other methods including the method of triads to derive the semantic space and obtained remarkably similar results in each case.

Henley's work represents one of several current approaches to the problem of specifying the form of the memory structure and the method of deriving the structure from similarity judgments. Most of the work to be discussed in this paper depends on a similar set of assumptions. If we accept the view that at least portions of the structure of semantic memory can be represented as a multidimensional space, then in order to specify any particular reasoning process we need only specify the appropriate algorithm operating on this structure. The case of similarity judgments is a particularly simple one because distance is a particularly

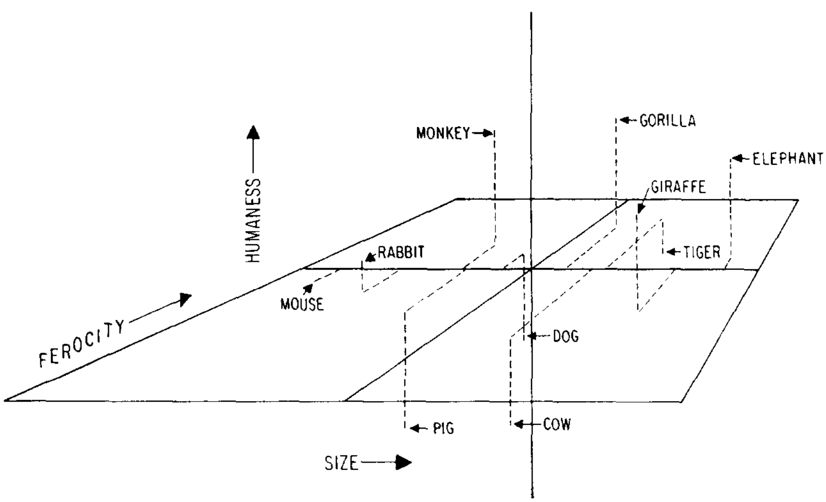

FIG. 1. The placements of a selected set of animals based on data from Henley (1969).

simple computation within the framework of a multidimensional Euclidean space.

A little thought will show that analogical reasoning can itself be considered a kind of similarity judgment in which not only the magnitude of the distance but also the direction must be indicated. Consider for example the assertion, A is to B as C is to D, the classic analogic paradigm. When we make this statement, we are simply asserting that the concept A is similar to concept B in exactly the same way and to exactly the same degree that concept C is similar to concept D. That is, in the multidimensional representation, we are asserting that the directed or vector distance between A and B is exactly the same as the directed or vector distance between C and D. Thus, in an analogy problem of the form A is to B as C is to what, the proper answer must be that concept which is most nearly the same vector distance from C as B is from A. These ideas are stated more formally and specifically in the set of assumptions listed below.

Consider an analogy problem of the form A:B::C:$(X_1, X_2, \ldots, X_n)$ (to be read A is to B as C is to which of the following: $X_1, X_2, \ldots,$ or X_n). It is assumed that:

A1. Corresponding to each element of the analogy problem there is a point in an m-dimensional space. (We denote, for example, the point corresponding to element A of the problem as A and say that the coordinates of A are the ordered sequence $\{a_j\}_{j=1,m}$, where a_j is the coordinate value of A on dimension j.)

A2. For any analogy problem of the form A:B::C:?, there exists a concept I such that A:B::C:I and an ideal analogy point, denoted I such that I is located the same vector distance from C as B is from A. The coordinates of I are given by the ordered sequence $\{c_j + b_j - a_j\}_{j=1,m}$.

A3. The probability that any given alternative X_i is chosen as the best analogy solution from the set of alternatives $X_1 \ldots, X_n$ is a monotonic decreasing function of the absolute value of the distance between the point X_i and the point I, denoted $|X_i - I|$.

To summarize, we assume that each element in an analogy can be represented as a point in an m-dimensional Euclidean space, that the ideal analogy solution is given by that point in the space which lies the same vector distance from C as B lies from A, and that, the closer a given alternative is to the ideal analogy solution, the higher the probability it will be chosen as the best analogy.

The intuitions behind these assumptions can be further illustrated with reference to the eight kinship terms in Fig. 2. Each term corresponds

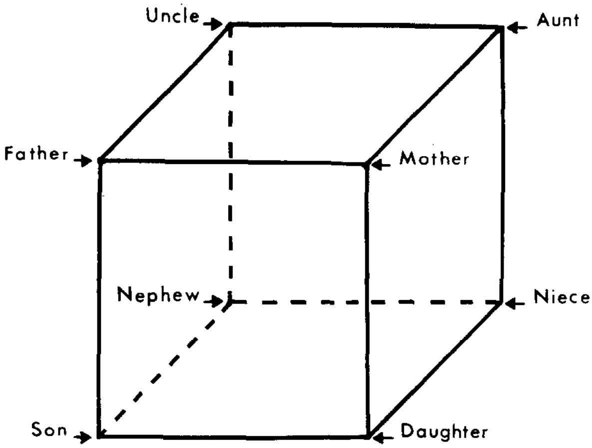

FIG. 2. Three-dimensional representation of relations among eight kinship terms.

to a corner of the three-dimensional similarity space shown in the figure. The coordinates of each term are shown beside it. Consider, as an example analogy: Son: Daughter: : Uncle: (Father, Nephew, Mother, Niece, Aunt). For simplicity, assume that the space shown in Fig. 2 satisfies A1. Then, under the assumption that the coordinates of Son are (0,0,0) and those of Aunt are (1,1,1), from A2 we get the coordinates of I to be

$$\{d_j + e_j - a_j\}_{j=1,3} = (0 + 1 - 0,\ 0 + 1 - 0,\ 1 + 0 - 0) = (1,1,1)$$

We observe that the term corresponding to the ideal solution for this analogy is Aunt, with coordinates (1,1,1). Thus, from A3 we conclude that Aunt should be the most probably chosen response. A second analogy is similar: Aunt: Nephew: : Mother: (Son, Father, Uncle, Daughter, Niece). Here, Son with coordinates (0,0,0) corresponds to the ideal analogy solution. The predictions correspond fairly well with our intuitive judgment of the best solutions.

The multidimensional space used here was a particularly simple one, composed of three binary dimensions which were assigned *a priori*, and the test was merely intuitive. We now turn to a more serious test of our assumptions.

EXPERIMENT I[2]

Assumption A1 of the theory requires a set of concepts located in a multidimensional space. The set chosen and used throughout Expts

[2] Experiment I was carried out by Sharon Wilson as part of an undergraduate research project under the sponsorship of the first author.

I, II, and III is the animal set scaled by Henley (1969) and described above. This set was chosen because the space seemed robust, because college student subjects are generally familiar with the set of words, and because we have Henley's experiments with which to compare our results. All analogy problems in Experiment I were of the familiar form $A:B::C:(D_1,D_2,D_3,D_4)$.

Procedure

Analogy problems were generated in the following way. Animal terms from the set of thirty were chosen at random without replacement. The first term chosen was the first element of the first analogy problem. The second chosen was the second element of the first analogy problem. The third chosen was the third element of the first problem. The fourth chosen was the first element of the second analogy problem, etc. This procedure was continued until the set of animal terms was exhausted (i.e., ten analogies were formed). All terms were then replaced in the pool of possible terms, and the procedure was continued until thirty unique analogy problems were formed. For each of the thirty analogy problems, four alternatives were chosen. One alternative was chosen

TABLE 1
Example Analogies and Solutions from Example I

	RAT:PIG::GOAT:——		CAMEL:DONKEY::RABBIT:——
	A. CHIMPANZEE B. COW C. RABBIT D. SHEEP		A. ANTELOPE B. BEAVER C. CAT D. TIGER
RANKS	1. B 2. D 3. C 4. A	RANKS	1. B 2. C 3. A 4. D
	FOX:HORSE::CHIPMUNK:——		LION:WOLF::GOAT:——
	A. ANTELOPE B. DONKEY C. ELEPHANT D. WOLF		A. CAT B. CHIMPANZEE C. GORILLA D. PIG
RANKS	1. C 2. A 3. B 4. D	RANKS	1. D 2. A 3. B 4. C

at random from among those animal terms within .5 units of the ideal analogy solution (.5 units corresponds roughly to the distance between CAMEL and ANTELOPE in Henley's scaling). If no animals fell within .5 units, the analogy was discarded and a new analogy was formed. After the first alternative was chosen, a second was chosen from among those animals between .5 and 1.0 units from the ideal solution. (One unit corresponds to the distance between DEER and ELEPHANT.) The third alternative was chosen among those animals from between 1.0 and 1.5 units from the ideal solution. (A distance of 1.5 units corresponds to the distance between ELEPHANT and PIG.) The fourth and final alternative was chosen from among those animals more than 1.5 units from the ideal solution.

Thirty-five subjects were recruited from a lower division psychology class at the University of California, San Diego. Subjects were run in two groups, one of 15 and one of 20 students. Each subject was given a five-page mimeographed booklet. The first page contained the instructions and an example. Subsequent pages contained the thirty analogy problems. The four response alternatives were listed in alphabetical order. Examples of the analogy problems are given in Table 1.

The directions were as follows:

DIRECTIONS:

An analogy task is one in which you recognize relationships of things or ideas to other things or ideas. In this task you will be given, the last term in the analogy will be missing. You will have four choices from which you will supply the term which seems to you to be the most analogous. Then you should indicate your second, third, and fourth choices, also. After that, you should indicate which of the terms you could consider to be analogous and which ones you think are not analogous by drawing a cut-off line between the two "groups."

Example: apple:tree::grape: ________

A. bush
B. vine
C. barrel
D. ground

1. B
___
2. D
3. A
4. C

The ordering of the relationship in an analogy is important. For instance, in the above example apple:tree::vine:grape would be incorrect because the relationship of apple *to* tree is not the same as that of vine *to* grape.

All of the terms in the following analogies are animal names. Feel free to go back and change any answer.

TABLE 2
Subjects' Rankings as a Function of Alternative Distance

Rank distance of the alternative from I	Subject-assigned ranks			
	1	2	3	4
1	.709	.180	.069	.046
2	.177	.546	.137	.129
3	.086	.160	.526	.226
4	.043	.111	.243	.600

Results and Discussion

The basic results of Expt I are shown in Table 2. Table 2 gives the proportion of responses, averaged over subjects and analogy problems, for which Rank 1 was given to the jth closest response alternative. Thus, the upper left hand entry of the table implies that 70.9% of the Rank 1 responses were given to the response alternative closest to the ideal point. Only 4.6% of the Rank 1 responses were given to the most distant response alternative.

Although the theory outlined in assumptions A1–A3 made no specific prediction beyond the Rank 1 data, it is clear that the entire table is consistent with the ideas behind the theory. As it stands, assumption A3 asserts only that a monotonic decrease in probability should be observed in the first column of Table 2. Before more specific predictions can be made, we require a more specific decision rule. The rule we propose is that developed by Luce (1959). We thus substitute for assumption A3 assumptions A3′ and A4 outlined below.

A3′. The probability that any given alternative X_i is chosen from the set of alternatives $X_1, \ldots X_n$ is given by

$$Pr(X_i|X_1, \ldots, X_n) = p_i = v(d_i) \Big/ \Big[\sum_j^n v(d_j) \Big],$$

where $d_i = |X_i - I|$ denotes the absolute value of the distance between X_i and I, and $v(\ )$ is a monotonically decreasing function of its argument.

A4. $v(x) = \exp(-\alpha x)$, where x and α are positive numbers.

Thus, A3′ is simply a restatement of Luce's choice rule, and A4 is an assertion that the monotonically decreasing function of A3′ is an exponential. The exponential was chosen for a number of reasons. It was chosen in part because Shepard (1957) found a good fit to an exponential generalization function over a similarly derived "psychological space" and in part because it is a simple one parameter decay function.

Assumptions A3′ and A4 were stated separately because of the obvious possibility that A3′ be correct and A4 false. Stated in this way, it is easy to separate the Luce choice rule and the exponential decay assumption.

Figure 3 is a plot of the predicted versus the observed number of subjects, ranking each of the four alternatives of each of the 30 analogy problems as the best solution to the analogy. The single parameter of the fit, α, was estimated by a least squares procedure to be 2.9. A product moment correlation computed from these data show a correlation of .933 between the predicted and observed values.

It appears from these data that the set of assumptions A1, A2, A3′, and A4 yield an adequate account of the first choice data. It is furthermore clear from the orderly nature of the data in Table 2 that a slight extension of the theory to include a theory of rankings may well account quantitatively for the entire matrix of data displayed in the table. Given our use of Luce's choice axiom in A3′, it is natural to accept Luce's extension of his choice axiom to cover rankings. We thus add an additional assumption, A5, to our list.

A5. We assume that the subjects rank a set of alternatives by first choosing the Rank 1 element according to A3′ and, then, of the remaining alternatives, deciding which is superior by application of A3′ to the remaining set and assigning that Rank 2. This procedure is assumed to continue until all alternatives are ranked.

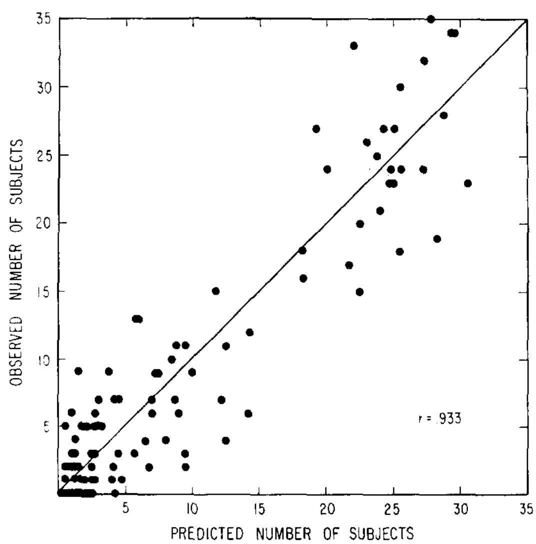

FIG. 3. Predicted versus observed number of subjects ranking each alternative as the best analogy solution.

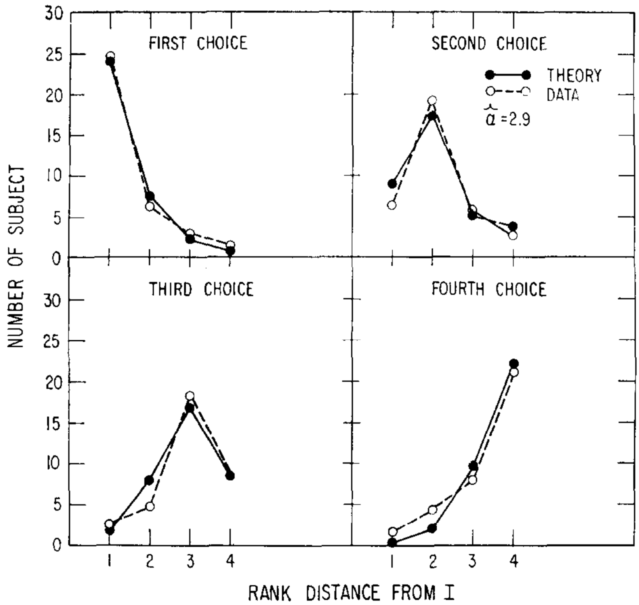

FIG. 4. Predicted versus observed mean number of subjects ranking each of the four response alternatives in each of the four ranks.

Figure 4 is a plot of the predicted and observed mean number of subjects, averaged over analogy problems, to assign rank j to the ith closest response alternative, for Ranks 1 through 4. The value of $\alpha = 2.9$ used to generate the theoretical curves was taken from fit of the first choice data discussed above. No new estimate was made to fit the second-, third-, and fourth-ranked alternatives.

Although the results of Expt I are certainly encouraging, it is somewhat unfortunate that the design of the experiment required an extension of the original three assumptions into five before quantitative tests of the theory could be made. The second experiment was designed to get a test of the basic theory unencumbered by the added response assumptions.

EXPERIMENT II

The goal of this second experiment was to extend the results of Experiment I by finding evidence bearing more directly on the basic assumptions of the model. The strong implication of assumptions A1–A3 is that the probability of choosing any particular alternative X_i as the best alternative depends on the ideal solution point I and on the alternative set $(X_1, \ldots, X_n)$, but not at all on the analogy itself. Thus, all

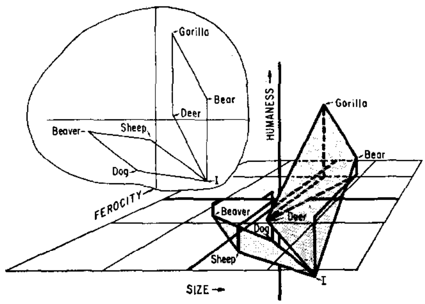

FIG. 5. Three-dimensional representation of two analogies with the same solution. The analogies are (a) GORILLA:DEER::BEAR:? and (b) BEAVER:SHEEP::DOG:?.

possible analogies with a given ideal solution point I and a given alternative set $(X_1, \ldots, X_n)$ should yield the same distribution of responses over the X_i. Experiment II was designed primarily to test this implication.

The logic of Experiment II is further illustrated in Fig. 5. The figure is a geometric representation of two analogies both with the same solution. The analogies are GORILLA:DEER::BEAR:? and and BEAVER: SHEEP::DOG:?. The response probabilities should depend on the distances of the various elements of the response set from I and not on which of the two analogies were presented.

Notice, incidentally, that an analogy is represented geometrically as a parallelogram in the plane passing through the three points corresponding to the three elements of the analogy. The insert shows the two analogies projected into the frontal plane.

Procedure

Twelve pairs of analogy problems of the type shown in Fig. 5 were constructed. Each pair had the same ideal point within a tolerance of .12 units (roughly the distance between a lion and a tiger). For each such analogy pair, two sets of alternatives (denoted set A and set B) were constructed. (Table 3 shows the analogy illustrated in the figure with its two alternative sets. Both sets are ranked in order of distance from I.) The alternative sets were constructed with the constraints that (1) there be no overlap (i.e., if an animal name appears in alternative set A, it does not appear in alternative set B; (2) the ith closest alternative for one set is roughly the same distance as the ith closest alternative for the other set, for all i; and (3) alternatives of a given set must be about equally separated in distance from one another by either .2, .25, .35, .4, or .45 units.

For each ideal point, the alternative sets were paired arbitrarily with

TABLE 3
Two Analogy Pairs with Alternative Sets from Experiment II[a]

Analogy pair 1	GORILLA:DEER::BEAR:___	BEAVER:SHEEP::DOG:___
Alternative set	A	B
	1. COW 2. PIG 3. TIGER 4. MONKEY	1. DONKEY 2. CAMEL 3. ELEPHANT 4. CHIMPANZEE
Analogy pair 4	CAT:SHEEP::LEOPARD:___	MOUSE:RACCOON::COW:___
Alternative set	A	B
	1. HORSE 2. DEER 3. FOX 4. RAT	1. ZEBRA 2. GIRAFFE 3. LION 4. CHIPMUNK

[a] Analogy pair 1 is also illustrated in Fig. 5.

the two analogies. The resulting 24 analogy problems were given to half of the 44 subjects. The 24 problems obtained by the opposite pairings were given to the other half. Hence, for each ideal point, a given subject encountered both members of the analogy pair and both alternative sets, but in only one of the two possible combinations.

The analogy problems were presented in booklet form with instructions similar to those of Experiment I. The analogy problems appeared in random order, and the alternatives for each analogy were also scrambled.

Results and Discussion

The theory holds that the response distribution should depend only on the ideal solution point and the alternative set, and not on the particular analogy problem. This prediction was tested by computing a χ^2 between the response distributions given to the two halves of each analogy pair for each response set. These χ^2 values are shown in Table 4. The summed χ^2 reaches a value of 109.7, with 60 degrees of freedom. This value is somewhat larger than would be expected by chance if the theory were true. On the other hand, a close look at the table will show that most of the deviation is contributed by a few comparisons. If, for example, the two largest values are ignored, the summed χ^2 reduces

TABLE 4

Analogy pair	Alternative set	χ^2	df[c]	Analogy pair	Alternative set	χ^2	df
1	A	1.83	2	7	A	7.33[b]	2
	B	.11	4		B	7.42[a]	2
2	A	.39	2	8	A	1.68	2
	B	.93	2		B	9.47	4
3	A	3.03	4	9	A	6.70[a]	2
	B	7.66	4		B	10.47[b]	2
4	A	2.47	2	10	A	.86	2
	B	3.39	4		B	.39	2
5	A	13.50[b]	2	11	A	1.47	2
	B	—	—		B	.12	2
6	A	25.71[b]	4	12	A	1.65	2
	B	2.12	4		B	1.00	2

Summed $\chi^2 = 109.7$, $df = 60$.
[a] Value exceeds .05 significance level.
[b] Value exceeds .01 significance level.
[c] In cases where the expected values of a cell was less than 5, cells were collapsed. Hence, the degrees of freedom in the various comparisons.

to 84.0 with 54 degrees of freedom, a relatively more probable value. Considering all of the possible sources of error, it seems safe to conclude, at least to a first order of approximation, that subjects' responses depend on the ideal analogy point and the alternative set, but not on the particular analogy problem. It should be pointed out, however, that in certain isolated examples, such as analogy pair 6, alternative set A, there is quite clearly an interaction between the particular alternatives and the analogy problem itself.

Since, on the whole, assumptions A1 through A3 are confirmed by our analysis, it makes sense to analyze these data with respect to the remaining assumptions of the theory. It is possible in the context of Experiment II to get a direct look at the form of $v(x)$ and thus a direct evaluation of assumption A4. To see how this is possible, consider assumption A3′. From A3′, we have

$$p_i = [v(d_i)] \Big/ \Big[\sum_j^4 v(d_j) \Big].$$

Taking the natural logarithm of each side yields

$$\ln p_i = \ln v(d_j) - \ln \Big[\sum_j^4 v(d_j) \Big].$$

Now, if $v(x) = \exp(-\alpha x)$ as asserted, we get

$$\ln p_i = -\alpha d_i - \ln \left[\sum_j^4 \exp(-\alpha d_j) \right].$$

In words, we find that the natural logarithm of the probability that alternative *i* is chosen is a linear function of alternative distance from *I* with slope $-\alpha$ and intercept $1 - n[\Sigma \exp(-\alpha d_j)]$. In Experiment II, alternative sets were designed such that several different analogy problems had the same set of interalternative distances, thus allowing averaging over analogy problems. There are only five different configurations of interalternative distances. Figure 6 shows the straight line fit of each of the five groups of analogies. The value of α was estimated to be 1.68. For the most part, the fit of the straight line is quite good. The only serious deviation is depicted in panel E. This is the condition in which the alternatives are closest together and also contains the lowest number of observations.

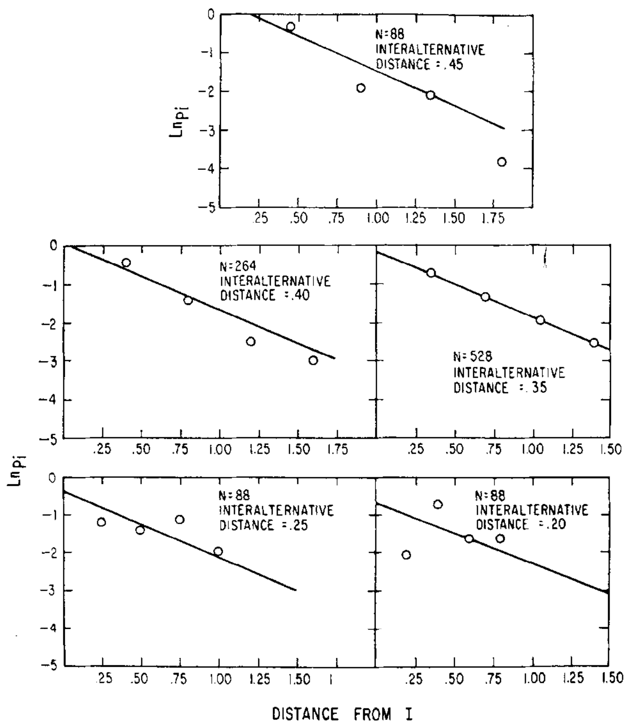

FIG. 6. Predicted and observed values of the natural logarithm of the probability that each of the four response alternatives will be chosen as the best analogy solution for each of five sets of interalternative distances.

Although in our analysis to this point we have looked only at the first-choice data, subjects in Experiment II, as those in Experiment I, were asked to rank the alternative sets. Figure 7 shows the predicted and observed mean number of subjects to rank alternative i in position j for each of the five groups of analogy problems. The value $\alpha = 1.68$ was taken from the fit of the first choice data discussed above. The figure

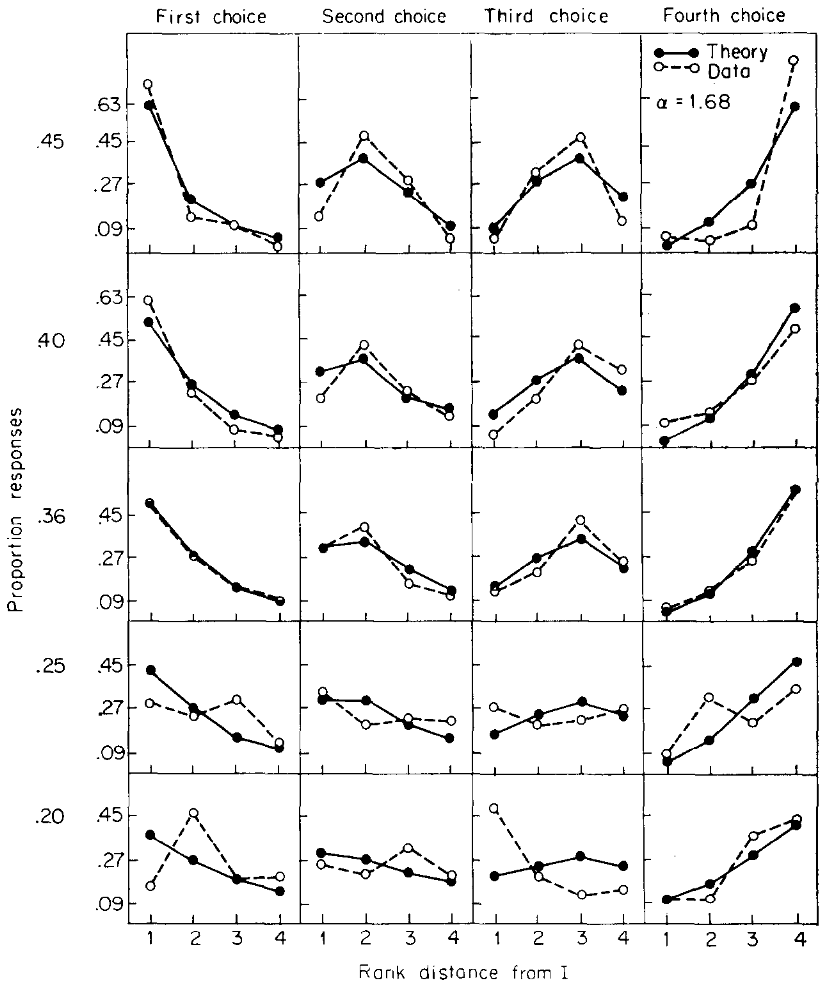

FIG. 7. Predicted versus observed mean number of subjects ranking each of the four response alternatives in each of the four ranks for five sets of analogy problems. The graphs are ordered from top to bottom in order of decreasing interalternative distances. These distances are given on the right hand side of each set of four graphs and vary from .45 to .20 units.

shows a qualitatively rather good fit between the predicted and observed values. Again, the case where the interalternative distances are the smallest shows the worst fit. The best fit occurs for the condition in which interalternative distances equals .35, a condition which incidentally has the greatest number of observations.

In summary then, Experiment II allowed a nonparametric test of assumptions A1 through A3. Although there were some exceptions, on the whole the results supported the model. In addition, Experiment II allowed direct investigation of the exponential function in A4. Again, the results supported the form of the model. Finally, a comparison of the entire set of rank data with predicted values lent more support to the entire set of assumptions A1–A5.

EXPERIMENT III

This research was originally conceived as an alternative approach to the problem of concept formation. It was felt that in natural settings much of concept formation involved learning by analogy, as opposed to instance generalization, the only form studied in the laboratory. Thus, Experiment III was designed to make use of the theory of analogical reasoning outlined above to facilitate the teaching of concepts.

Assumption A2 implies that each statement of the form A:B::C:? implies the existence of some concept[3] I against which alternative sets are compared to find the best alternative. The logic of Experiment III was to give a name to this implied concept I and then to see if subjects are able to use this new concept with a new name in the same way that they are able to use concepts that they already know. For example, the experimenter might assert that A:B::C:GOX where GOX is the concept to be learned. He would then ask the subject to use GOX in other cognitive task to see if the subject is able to manipulate it in the way he manipulates other concepts. The subject might be asked to solve analogies of the form A:B::GOX:$(X_1, \ldots, X_n)$, or to judge the similarities between GOXes and other animals, or to give a verbal description of a GOX. In short, we would expect that if the subject had actually understood the analogy A:B::C:GOX he would be able to do anything with the concept GOX that he could with any other concept—he would have formed a new concept.

Procedure

Three points were chosen in Henley's animal space. These points were labeled BOF, DAX, and ZUK and represented the three concepts

[3] This concept must be regarded as a potential concept because in general it hasn't been named in the language.

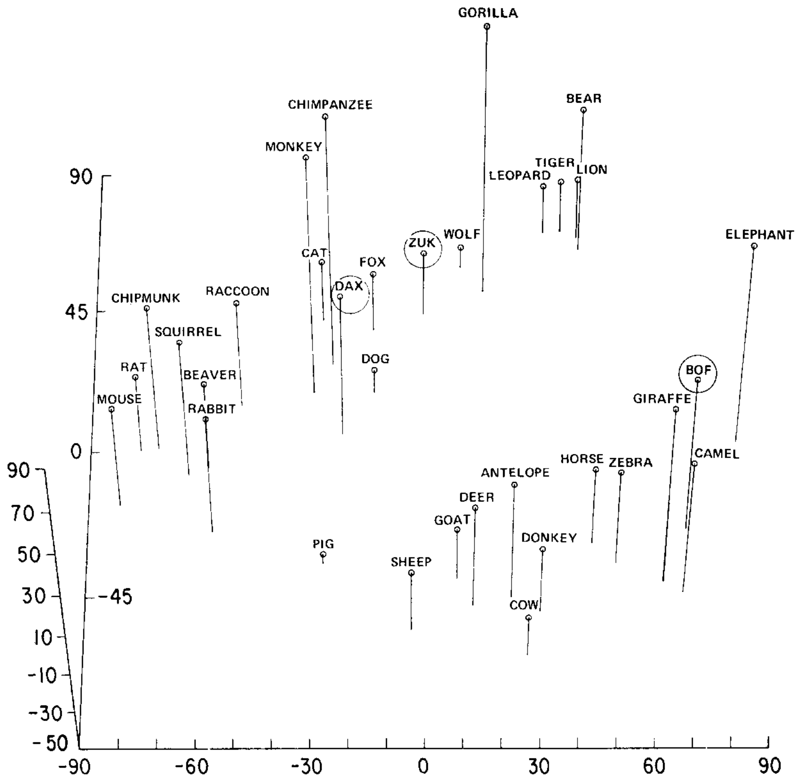

FIG. 8. Three-dimensional representation of Henley's animal space including the location of the three artificial animals, BOF, DAX, and ZUK.

that the subjects were to learn. The points were chosen such that they occupied fairly remote parts of the space. Figure 8 is a three-dimensional representation depicting the location of the three artificial animals, BOF, DAX, and ZUK, relative to the thirty mammals scaled by Henley. Thus, a BOF lies between an ELEPHANT and CAMEL, a DAX near CHIMPANZEE, and a ZUK between FOX and WOLF.

An anticipation method of teaching was employed. That is, subjects were first given an analogy problem of the form A:B::GOX:$(X_1, \ldots, X_4)$ and then were asked to make a guess as to the best alternative. Following their guess, subjects were informed as to the correct alternative along with a ranking of the remaining alternatives from best to worst. After a period of study, the subjects were given another analogy problem of the same form, and the process was repeated. Table 5 shows an example of one trial of the analogy learning task. Response alternatives were chosen with the restrictions that the closest alternative was within .12

TABLE 5
An Anticipation-Training Trial from Experiment III

View 1	1. SHEEP: CAT	
	BOF : ?	
	(a) Rabbit	
	(b) Chimpanzee	
	(c) Leopard	
	(d) Bear	
View 2	SHEEP: CAT	
	BOF : BEAR	
	BEST ANSWER:	(d) Bear
	Second	(c) Leopard
	Third	(b) Chimpanzee
	Third	(a) Rabbit

units of the ideal solution point, other alternatives roughly trisecting the distance from the point to the most distant alternative.

Applying the .12 unit tolerance, there were available in the space 20 analogies to teach BOF, 18 to teach DAX, and 17 to teach ZUK. Each of 25 subjects was given the training sequence on each of the three concepts. Following training, subjects were given rating sheets asking them to rate, on a scale from 1 (labeled "very similar") to 10 (labeled "very different"), the dissimilarity of each of the three artificial animals with each of the 30 animals in Henley's space and with each other. Following these ratings, subjects were asked to give a verbal description of each of the three animals.

We thus have three measures of the degree to which the concepts which we have attempted to teach have in fact been learned by the subjects. First, we can observe their ability to solve analogies while they are learning the concepts. We expect that subjects will respond at chance level on the first trial of each training sequence and, thereafter, will approach more or less quickly the response patterns of subjects in Experiments I and II. Secondly, we can compare our subjects' judgments of dissimilarity with those expected. Finally, we can look at their verbal descriptions to see the completeness of the concept and the extent to which different subjects describe the various artificial animals similarly.

Results and Discussion

Our first prediction regards the way in which these artificial animals are used to solve analogy problems. If the subjects have really learned the concepts of BOF, DAX, and ZUK, their behavior during the latter phases of the training sequences should be indistinguishable from the behavior

TABLE 6
Predicted and Observed Mean Number of Subjects Choosing the ith Closest Alternative as the Best[a]

Rank of alternative distance from I							
1		2		3		4	
Pred[b]	Obsd	Pred	Obsd	Pred	Obsd	Pred	Obsd
12.4	13.3	6.6	8.2	3.1	2.5	1.6	1.6

[a] Predicted values generated under the assumption that $\alpha = 1.68$.
[b] Pred, predicted; obsd, observed.

of our subjects in Experiments I and II. Since it appeared that learning was complete after, at most, five analogy anticipation trials, the anticipation data after the fifth trial should be the proper comparison. Table 6 shows the mean number of subjects choosing each of the four alternatives as the best. The average is taken over all anticipation trials after trial five on each of the three concept problems. The predicted values are the average across items of those expected based on the parameter value of $\alpha = 1.68$ taken from Experiment II. The closeness of fit of these observed and expected values are definitely consistent with the idea that the concepts BOF, DAX, and ZUK can be manipulated and used just as the other animal concepts in our other experiments.

In spite of the goodness of fit of these data, these comparisons are indirect. It would seem that the similarity judgments would yield a more direct measure of the extent to which we have actually succeeded in teaching the concepts we set out to teach. Since BOF, DAX, and ZUK are points exactly located in Henley's similarity space, we can make direct predictions of dissimilarity ratings by measuring the distance between each artificial animal and each of the other animals in the space. We predict, in a strictly nonparametric way, that mean dissimilarity ratings will increase monotonically with distance in the space. Figures 9, 10, and 11 are plots of our observed dissimilarity rating versus distance measured in Henley's space. The product moment correlations were $r = .95$, .90, and .92 for BOF, DAX, and ZUK, respectively. The Kendal's tau values were $\tau = .79$, .75, and .73 for the three animals.

Although on a priori basis correlations between .9 and .95 seem good, it is not clear how good the correlation should be in order to be comparable to real concepts. To help answer this question, Nancy Henley kindly supplied us with her raw data for comparison. We were thus able to compute similar correlation coefficients between her subjects' dissimilarity judgments and her derived semantic space. These coefficients were

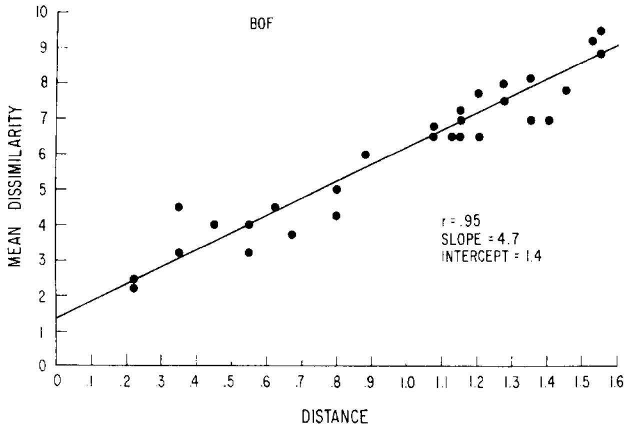

FIG. 9. Mean dissimilarity judgments between BOF and each of the 30 mammals as a function of the distance of the mammal from BOF.

found to vary between $r = .98$ for CAMEL, CHIPMUNK, and SQUIRREL, and $r = .86$ for GORILLA. The values of tau varied as low as $\tau = .58$ for BEAR. Furthermore, of the 30 animals, 11 had values of r higher than .95, and 6 had values lower than .90. The remaining 13 had values between .90 and .95. The conclusion thus seems clear. The dissimilarity judgments given by our subjects comparing BOF, DAX, and ZUK to each other and to each of the other 30 animals are every bit as predictable as are subjects' responses with real animal names.

One final computation of interest was carried out with the dissimilar-

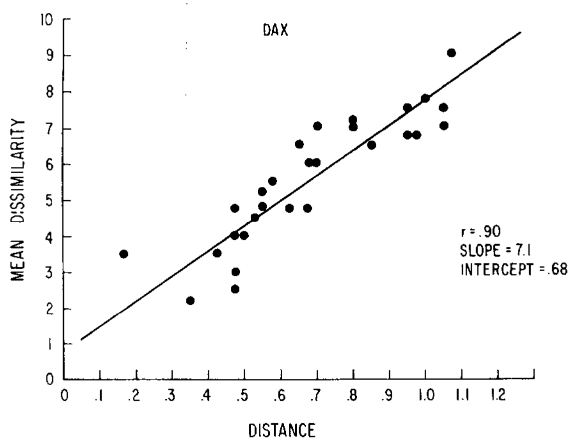

FIG. 10. Mean dissimilarity judgments between DAX and each of the 30 mammals as a function of the distance of the mammal from DAX.

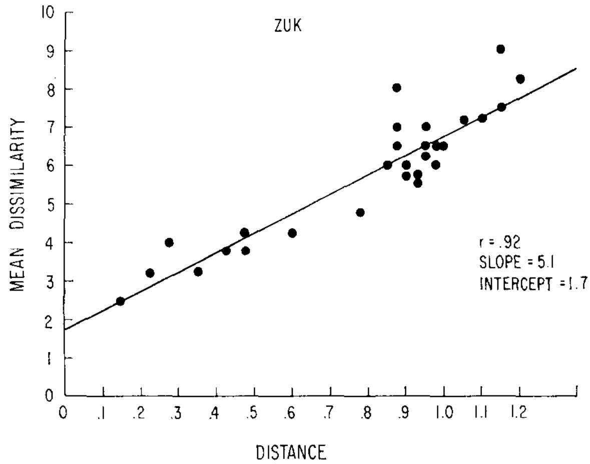

FIG. 11. Mean dissimilarity judgments between ZUK and each of the 30 mammals as a function of the distance of the mammal from ZUK.

ity data. We added our subjects' data to that of Henley, thus increasing the number of animals from 30 to 33. We then inserted the data into a version of TORSCA[4] to obtain a revised scaling. Figures 12 and 13 show a superposition of the revised scaling on Henley's original semantic space. The arrows show the migration of each of the points from the first scaling to the second. A value of Kruskal's index of stress of 10% was obtained in the rescaling as compared to a value of 9.4% in Henley's original scaling. Of special interest is the movement of the three artificial animals. The origin of the arrow indicates the intended location of the concepts as indicated by the rescaling, BOF moved a total of .153 units, DAX moved .277 units, and ZUK moved .132 units. The largest movement of the animal concepts occurred with ELEPHANT, which moved .094 units. The fact that the artificial animals moved more than the real animals should not be surprising on three counts. First, and most obviously, the vast majority of the data going into the rescaling was the same data which determined the location of the animals in the first place. Secondly, it should be recalled, the training analogies were correct with a tolerance of .12 units. Hence, our training procedure insures that we will teach what we want only within .12 units. Thirdly, although the scaling procedure requires that the same monotonic transformation be applied to all of the dissimilarity judgments in a given scaling, our procedure tended to promote context effects. That is, our subjects were given a sheet asking them to judge the dissimilarities of

[4] Thanks are due to Richard Meltzer who provided a copy of TORSCA and was otherwise instrumental in carrying out this analysis.

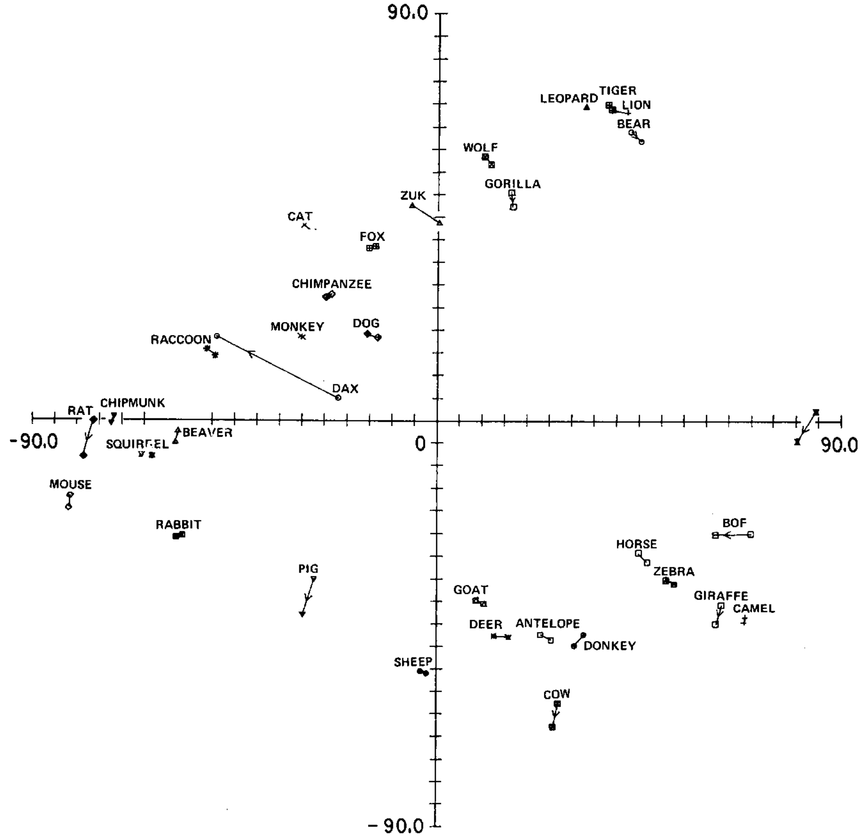

FIG. 12. A superposition of dimensions 1 and 2 ("size" and "ferocity") of Henley's original scaling and the scaling generated from including data from Experiment III. The arrows indicate change from original to revised scaling.

DAX and ANTELOPE, DAX and BEAR, etc., down to DAX and ZEBRA. Then a similar procedure was used for the BOF judgments, etc. It is quite likely that this systematic presentation of pairs to be judged results in different standards of judgment being used from one block of judgments to the next. A glance at Figure 9, 10 and 11 indicates the problem. Although the regression line fits very well in all three cases, the slope for DAX is about 50% higher than for BOF and ZUK. The reason seems to be that DAX is a relatively central animal, and thus there are no animals very distant from it. Nonetheless, the entire range, 1–10, is used. Hence, the higher slope.

The final datum from Experiment III is the verbal descriptions. It is never clear what should be made from such data; nonetheless, they

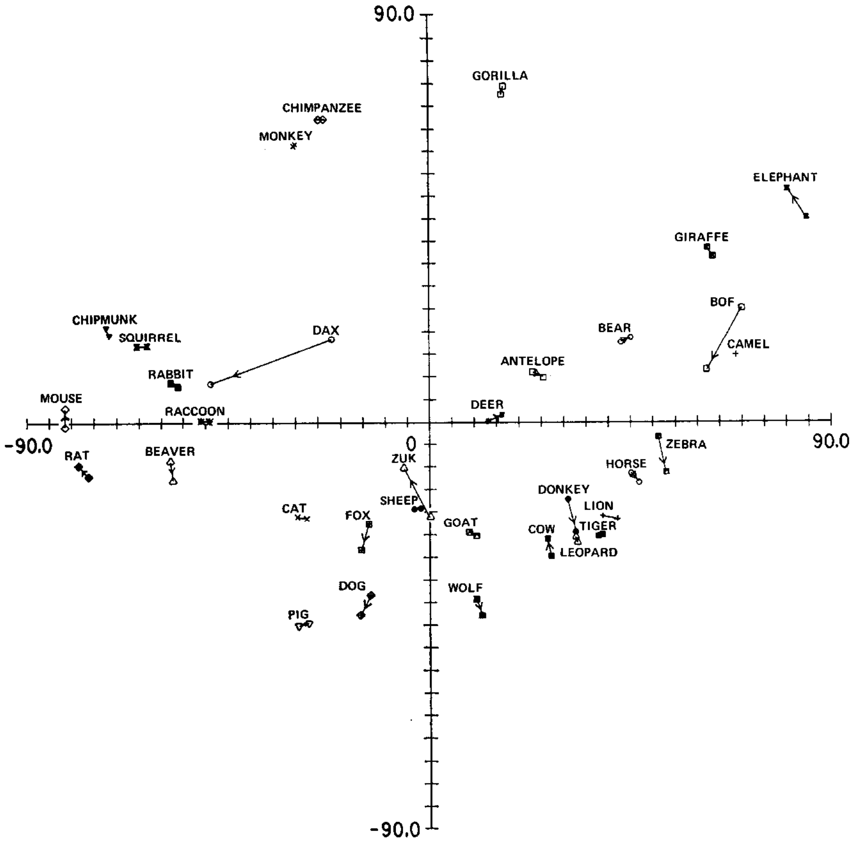

FIG. 13. A superposition of dimensions 1 and 3 ("size" and "humanness") of Henley's original scaling and the scaling generated from including data from Experiment III. The arrows indicate change from original to revised scaling.

occasionally make interesting reading. To that end, Tables 7, 8, and 9 give a compilation of subjects' comments regarding their ideas of the three animals.

GENERAL DISCUSSION AND CONCLUSIONS

We began our discussion with an analysis of reasoning. We suggested that reasoning is the collection of processes or algorithms which operates on organized memory in various information retrieval tasks. Any particular reasoning task, such as analogical reasoning, is the application of a particular type of data retrieval. The problem of characterizing a reasoning process is simply the specification of a data base from which information is retrieved according to a specific retrieval strategy or

TABLE 7
Verbal Description: BOF (18 Subjects)

SIZE:
- 5 Large
- 5 Camel-sized
- 3 Horse-sized
- 1 Rather large
- 1 Relatively large
- 1 Fairly big
- 1 Medium-sized mammal
- 1 Bigger than donkey
- 1 Very large antelope
- 1 Somewhat smaller than elephant
- 1 Larger than elephant

SHAPE:
- 1 Slender
- 1 Heavy
- 1 Visualize as big donkey

APPENDAGES:
- 1 Antlers or horns
- 1 Probably has horns
- 1 Small horns
- 1 Horns or humps

LEGS:
- 5 Four legs
- 1 Uses all four legs
- 2 Long legs
- 2 Hooved feet

NECK:
- 2 Long
- 1 Extended but shorter than giraffe
- 1 Shorter than giraffe

HAIR:
- 2 Short
- 2 Fur
- 1 Rather long (unlike camel, giraffe, cow)

WILDNESS:
- 2 Wild
- 1 Completely domesticated
- 1 Could be used as work animal

MOVEMENT:
- 1 Swift-moving
- 1 Runs very fast (like antelope)
- 1 Much more agile than camel
- 1 Not as coordinated as giraffe
- 1 Doesn't move very swiftly
- 1 Strong (like elephant or camel)

LIVE:
- 1 In desert or jungle
- 1 On plains

FOOD:
- 1 Non-carnivore
- 1 Probably eats plants
- 1 Grazes
- 1 Naturally eats grain and leaves

TYPE:
- 1 Mammal

INTELLIGENCE:
- 1 Not as intelligent as giraffe

LIFE STYLE:
- 1 Travels in herds with leader
- 1 Fairly mobile (like giraffe or zebra)

SIMILAR TO:
- 6 Camel
  - 1 Features similar to camel
  - 1 Camel
  - 1 Cross between camel and horse
  - 1 Somewhere between giraffe and camel
  - 1 Similar to camel or giraffe, closer to giraffe
  - 1 Cross between camel, giraffe, zebra, elephant
- 5 Giraffe
  - 1 Like giraffe but shorter neck (mentioned camel first)
  - 1 Somewhere between giraffe and camel
  - 1 Similar to camel or giraffe, closer to giraffe
  - 1 Very similar to domesticated giraffe
  - 1 Cross between camel, giraffe, zebra, elephant
- 2 Goat
  - 1 Cross between goat and deer but camel-sized
  - 1 Something like goat or sheep
- 2 Antelope-Deer
  - 1 Very large antelope
  - 1 Cross between goat and deer but camel-sized
  - 1 Cross between camel and *horse*
  - 1 Visualize as big *donkey*
  - 1 Something like goat or *sheep*
  - 1 Cross between camel, giraffe, *zebra*, elephant
- 2 Gorilla
  - 1 Characteristics similar to gorilla
  - 1 Possibly slight similarities to a primate

TABLE 8
Verbal description: DAX (18 Subjects)

SIZE:
- 6 Small
- 2 Medium-sized
- 1 Rather small
- 1 Very large
- 4 Size of beaver
- 1 Larger than beaver
- 2 Size of rabbit
- 1 Somewhat larger than rabbit
- 1 Size of squirre
- 1 Size of raccoon
- 1 2-ft-long
- 1 20–30 pounds

LEGS:
- 2 Four legs
- 3 Can stand on two legs
  - 1 Can support self on two as well as four
  - 1 Can manipulate with front feet while standing on rear
  - 1 Can stand on hind legs
- 1 Short legs
- 1 Has claws

APPENDAGES:
- 1 Long bushy tail
- 1 Long husky tail
- 1 Short tail

HAIR:
- 4 Furry
- 1 Long fur
- 2 Hairy (1 Like chimpanzee)

FEATURES:
- 1 Small ears, bright eyes, all that
- 1 Large front teeth
- 1 Big ears
- 1 Features of rabbit
- 1 Obviously green

WILDNESS:
- 1 Possibly domesticated
- 1 Close to being domestic; lives around man (like mouse)
- 1 Can be seen in wild or kept as pet (like rabbit)

MOVEMENT:
- 1 Fairly quick
- 1 Fast-moving
- 1 Quick moving, agile
- 1 Can climb trees well (like raccoon)

LIVE:
- 2 In forest (perhaps trees); doesn't like water as does beaver
- 2 Near water
  - 1 Often near water
  - 1 Dexterous in water, like beaver

FOOD:
- 1 Herbivorous or possibly omniverous
- 1 East young plant shoots

TYPE:
- 5 Rodent family
- 1 Mammal

INTELLIGENCE:
- 1 Clever
- 1 Dumb

PERSONAL CHARACTERISTICS:
- 1 Industrious
- 1 Cuddly
- 1 Fairly timid, not aggressive

SIMILAR TO:
- 4 Rabbit
  - 2 Larger than rabbit, however
  - 1 Like rabbit in life and size
  - 1 Like rabbit
- 2 Beaver
  - 1 Larger than beaver, however
  - 1 Cross between beaver, chimp, raccoon
- 2 Raccoon
  - 1 Very similar to squirrel and resembles raccoon
  - 1 Cross between beaver, chimp, raccoon
- 1 Squirrel
  - 1 Very similar to squirrel and resembles raccoon
- 1 Chimpanzee
  - 1 Cross between beaver, chimp, raccoon

TABLE 9
Verbal Description: ZUK (19 Subjects)

SIZE:
- 3 Small
- 3 Medium
- 1 Medium small
- 1 Largish
- 2 Size of fox
- 1 Same or smaller than fox
- 1 Size of large fox; smaller than donkey
- 1 Larger than fox; smaller than lion
- 1 Size of dog
- 1 Larger than dog
- 1 Slightly smaller than wolf
- 1 Size of raccoon or beaver, like DAX

SHAPE:
- 1 Similar in proportion to dog
- 1 More like dog or cat than zebra in build and fur; maybe like a great big mouse

HAIR:
- 2 Furry
- 1 Moderately hairy
- 1 Probably has shaggy fur (longer than fox)
- 1 Long-haired coat like goat and wolf
- 1 More like dog or cat than zebra in fur

APPENDAGES:
- 1 Tail
- 1 No horns; long flat tail

LEGS:
- 1 Four legs
- 1 Four stubby legs

FEATURES:
- 1 Similar to dog in features
- 1 Probably has unusual colorings
- 1 Large front teeth

WILDNESS:
- 2 Wild
- 1 Fierce
- 1 Attacks smaller animals
- 1 Clearly not domesticated; less wild than fox
- 1 Fairly aggressive, or at least mischievous and prying

MOVEMENT:
- 1 Quick
- 1 Fast runner
- 1 Very fleet, swift-moving

MOVEMENT:
- 1 Not as fleet as fox
- 1 Slow
- 1 Agile like wolf

LIVE:
- 1 Probably in forest
- 1 In woods; in ground or caves or hollowed trees; uses leftover tunnels
- 1 Perhaps in large holes in ground
- 1 Excellent on mountains, like goat

LIFE STYLE:
- 1 Not in packs like wolf; good family life

FOOD:
- 4 Predatory
  - 1 Seeks prey (like cat)
  - 1 Attacks smaller animals
  - 1 Preys on chickens and sometimes sheep, goats; problem to farmer
- 2 Carniverous

TYPE:
- 1 Great cat family
- 1 Probably canine; possibly feline
- 1 Dog family
- 1 Mammal

INTELLIGENCE:
- 1 Clever
- 1 Not as intelligent as fox

SIMILAR TO:
- 5 Cat family
  - 1 Cat-like (size largish)
  - 1 More like dog or cat than zebra in build or fur
  - 1 Cross between cat and canine families
  - 1 Like lion or other large cat
  - 1 Maybe is cheetah
- 4 Canine family
  - 1 Like dog
  - 1 More like dog or cat than zebra
  - 1 Cross between cat and canine families
  - 1 Like wild dog
  - 1 Cross between goat and wolf
- 3 Fox
  - 1 Resembles fox
  - 1 Cross between fox and dog
  - 1 Most like fox or any wild dog
- 2 Farm Animals
  - 1 Cross between *goat* and wolf
  - 1 Horse-like
- 1 Maybe like a great big *mouse*

algorithm. We chose to characterize the data base for our theory of analogical reasoning as a multidimensional Euclidean space. Given that specification of the data base, our choice for a reasoning algorithm becomes very natural. It would appear that the results of Experiments I, II, and III confirm the accuracy of our description of analogical reasoning among animal concepts. There are many possible ways to generalize the results we have found. The most straightforward of these would be to search for other sets of concepts which are well embedded in a multidimensional similarity space and show that our theory also gives an accurate description for analogy formation among these concepts. We have, in fact, tried some other spaces. In particular, we have generated analogies from the Munsell color space and have found essentially similar results.[5] Problems arise, however, when relations among words are not well represented in a multidimensional space. (Hierarchial relations are such examples. Where, for instance, would the word ANIMAL fit in Henley's space?) These kinds of examples seem to demand a more general representation of memory structure which is capable of handling more complex sorts of relationships while still holding the multidimensional representation as a special case. It is interesting in this regard that Quillian's (1969) Teachable Language Comprehender (TLC) seems to be such a generalization. Lexical information is encoded in TLC as a class name (superset) with several specializing modifiers (properties).[6] The properties themselves are written as attribute-value pairs. Thus, ORANGE might be defined as a FRUIT (superset) with properties COLOR (attribute)-ORANGE (value), SIZE-MEDIUM, and TASTE-SWEET; whereas LEMON might be defined as a FRUIT with COLOR-YELLOW, SIZE-SMALL, and TASTE-SOUR. If Quillian encoded a set of words, all with the same superset, and which differed from one another only with regard to values on a common set of attributes, the words could be equivalently encoded with a multidimensional representation. It is perhaps not coincidental that the set of words Henley scaled have exactly these characteristics.

In conclusion then, it should be clear that we are not suggesting that a multidimensional representation is sufficiently rich to encode all types of semantic relationships. Therefore, we are not proposing that this

[5] This experiment was carried out by Glenn Rice as part of a research project under the sponsorship of the first author.

[6] Although, in the illustration of TLC, we use only traditional sorts of attribute-value pairs in which attributes are dimensions and values are simple values on dimensions. Quillian's general model allows both to be more general. Thus, he might define CANARY as a BIRD with COLOR-YELLOW, SING-(SONG-PRETTY), where the verb sing is considered an attribute, and the value is itself an attribute-value pair.

model of analogical reasoning is completely general. It seems rather that the semantic relationships among certain sets of concepts can be represented by a multidimensional structure. Furthermore, whatever algorithm subjects actually employ in generating answers to verbal analogy problems seems to be isomorphic to our "parallelogram" rule. It is, of course, not satisfying to stop with a model of semantic relationships with such a limited range of generality. We are, in fact, currently developing a more complex network representation of semantic information (cf., Rumelhart, Lindsay, and Norman, 1972) which might serve as a more general representation on which we can propose and test more general models of analogical reasoning.

REFERENCES

BARTLETT, SIR FREDERIC. *Thinking: An Experimental and Social Study.* New York: Basic Books, 1958.

BECKER, J. D. The modeling of simple analogic and inductive processes in a semantic memory system. *Proceedings of the International Joint Conference on Artificial Intelligence.* Washington, D.C.: 1969. Pp. 655–668.

EVANS, T. G. A heuristic program to solve analogy problems. *Proceedings of the Spring Joint Computer Conference, AFIPS,* 1964, **28**, 327–338.

HENLEY, N. M. A psychological study of the semantics of animal terms, *Journal of Verbal Learning and Verbal Behavior,* 1969, **8**, 176–184.

HUNT, E. Computer simulation: artificial intelligence studies and their relevance to psychology. *Annual Review of Psychology,* 1968, **19**, 135–168.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrica,* 1967, **29**, 1–27.

LUCE, R. D. *Individual Choice Behavior.* New York: Wiley, 1959.

MILLER, W. S. Manual of the Miller Analogies Test. New York: The Psychological Corporation, 1960.

MINSKY, M. Artificial intelligence. *Scientific American,* 1966, **215**, 246–263.

OPPENHEIMER, ROBERT. Analogy in science. *American Psychologist,* 1956, **11**, 127–135.

POLLYA, G. *How to Solve It.* Garden City, New York: Doubleday Archer, 1957.

QUILLIAN, M. R. The teachable language comprehender. *Communications of the Association for Computing Machinery,* 1969, **12**, 450–476.

REITMAN, W. *Cognition and Thought.* New York: Wiley, 1965.

RUMELHART, D. E., LINDSAY, P. H., NORMAN, D. A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), *Organization of Memory.* New York: Academic Press, 1972.

SHEPARD, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika,* 1957, **22**, 325–345.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrika,* 1962, **27**, 125–140.

YOUNG, F. W., & TORGERSON, W. S. TORSCA, A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science,* 1967, **12**, 498.

(Accepted August 28, 1972)