

Judging Similarity Among Strings Described By Hierarchical Trees

Lola L. Lopes

University of Wisconsin, USA

Michael D. Johnson

University of Chicago, USA

The paper compares the tree-theoretical model of similarity judgment (in which the similarity between two objects is a function of the distance between them in a conceptual tree) with an averaging model of similarity judgment that is drawn jointly from information integration theory and from current research indicating the prevalence of anchoring and adjustment mechanisms in judgment. Results of an experiment are presented that suggest that even when subjects organize conceptual material as a hierarchical tree, judgments of similarity among the objects are better accounted for by an averaging mechanism than by distances in the tree. These data are discussed in terms of the differences between the representation in which knowledge is encoded and the processes that operate on the represented information.

Psychological treatments of similarity are dominated by structural theories of how concepts are represented in the mind, with competing systems centered around formalisms as diverse as coordinate spaces (Kruskal 1964a, b; Shepard 1962a, b), tree structures (Cunningham 1978; Cunningham and Shepard 1974; Johnson 1967), and sets of discrete features (Shepard and Arabie 1979; Tversky 1977). This emphasis on representation is not merely cosmetic: it reflects the basic, but often unvoiced, assumption that similarity is implicit in cognitive structure. In many ways this assumption captures the intuition that confusable or substitutable objects or concepts tend to share many characteristics. But the assumption is dangerous scientifically since it effectively removes from consideration any empirical examination of how commonalities among stimulus objects influence behavior, and replaces this with logically appropriate, but empirically untested, normative schemes for quantifying similarity.

The goal of the present paper is to bypass consideration of similarity as an abstract conceptual entity, and to consider instead the psychological process of judging similarity. Our exposition has three parts: first, we describe an averaging model of similarity judgment that has had some previous success in accounting for judgments of similarity among kinship terms (Lopes and Oden 1980); second, we show how the averaging model can be applied to judgments of similarity among stimuli described by hierarchical trees, and contrast this with the approach taken by current tree-theoretic models of similarity; third, we present the results of an empirical test of the relative ability of the two models to account for judgments of similarity among strings generated from, and described by, hierarchical trees.

Averaging and adjustment in similarity judgment

The averaging model of similarity judgment draws both from Anderson's (1974) information integration theory and from Tversky and Kahneman's (1974) research on judgmental heuristics, with the former providing the quantitative framework and the latter providing the cognitive mechanism. Equation 1 describes the averaging model algebraically. The degree of similarity between stimuli, $s(a,b)$, is assumed to vary inversely with a weighted average of the componential differences between the stimuli on each of a set of relevant dimensions [1] or attributes:

$$s(a, b) = \alpha - \frac{\sum_{i=1}^n w_{iab} d_{iab}}{\sum_{i=1}^n w_{iab}} \quad (1)$$

where d_{iab} is the scale difference between stimulus a and stimulus b on the i th dimension or attribute, w_{iab} is the associated weight, and α is a scaling constant that expresses averaged differences in terms of similarity and that allows for left/right biases in the use of the response scale.

It is important to stress that equation (1) is not intended to describe the *process* of judging similarity. We believe instead that similarity judgments are produced through application of cognitive anchoring and adjustment (Tversky and Kahneman 1974), with averaging emerging as the *consequence* rather than the goal of the adjustment strategy. In the general case we propose that subjects produce similarity judgments by “anchoring” their judgment at the similarity value (d_{iab}) associated with some salient dimension, and then adjusting that value in accord with information from other dimensions so that after each adjustment the “new” value of the similarity judgment lies in the interval between the “old” value and the value of the dimension just considered. The weights in equation (1) are assumed to derive from the order in which the differences on the various dimensions are processed, and this order, in turn, is assumed to reflect the relative salience of the various similarity values. Thus, salient values (e.g., those that are associated with important dimensions or those that are relatively extreme in the set of potential values) will be weighted relatively more heavily than non-salient values both since they benefit from judgmental primacy and since they call for larger adjustments.

Similarity among hierarchical stimulus strings

The purpose of the present research was to test the averaging model of similarity judgment against a tree-theoretic distance model (e.g., Cunningham 1978; Cunningham and Shepard 1974) using stimuli that could be represented as nodes in a hierarchical tree. The stimulus set we used is the set of eight number strings shown in fig. 1 plus the analogous set of eight letter strings obtained by replacing the numerals 1, 2, 3, and 4 by the letters a, b, c, and d, respectively. The strings were generated by

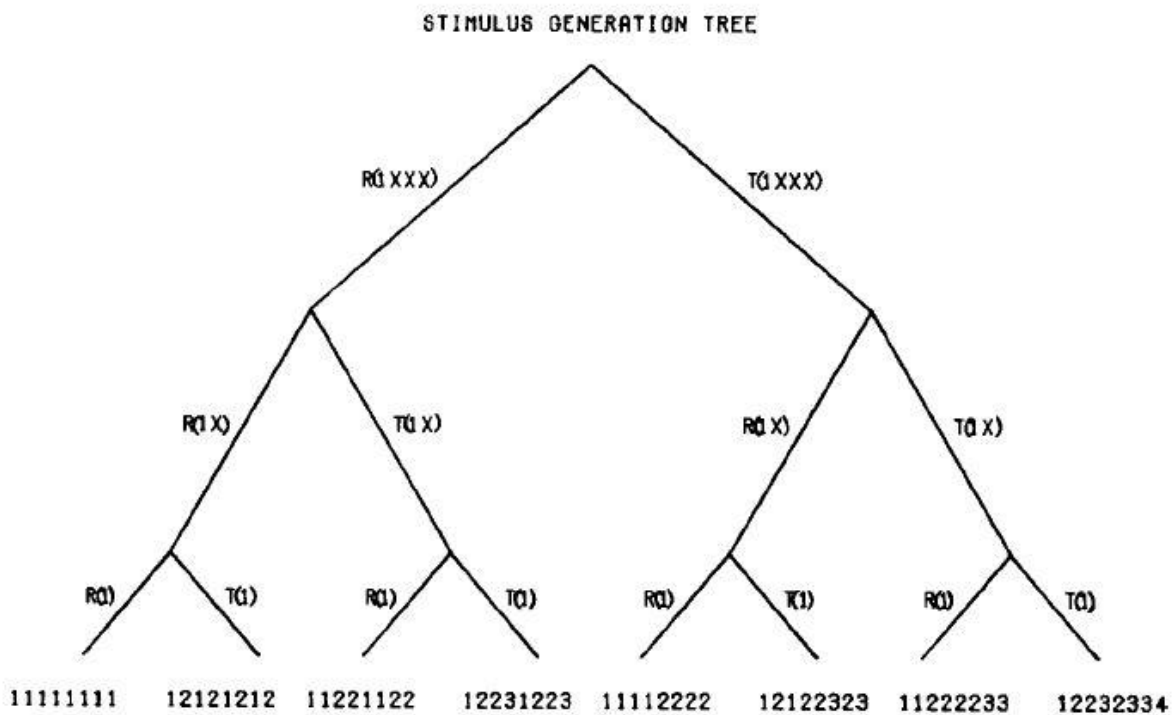


Fig. 1. Tree showing hierarchical structure of stimulus set.

beginning either with a “1” or an “a” and then doubling the length of the string three times by successive applications of string repetition and/or string transposition (Restle 1970). In repetition, a string is doubled by repeating it exactly so that, e.g., “1” becomes “11”, “12” becomes “1212”, “1223” becomes “12231223”, and so forth. In transposition a string is doubled by incrementing the original symbols by one place and then appending the incremented string to the original so that, e.g., “1” becomes “12”, “12” becomes “1223”, “1223” becomes “12232334”, and so forth.

Fig. 1 is a stimulus tree, which is to say that it represents the underlying design of the stimulus set. Whether or not it also represents the way any or all subjects would organize the stimuli conceptually is another issue—one that is best settled empirically. But if we assume for the moment that the figure fairly represents the way in which at least a few subjects would organize the stimuli, then it is possible to show how the averaging model and the distance model differ in their accounts of how such trees function in the similarity judgment process.

Averaging model

In the averaging model, a subject judges similarity by a stepwise process of anchoring and adjustment in which a series of unidimensional judgments are integrated to produce an overall value. For example, given the strings 11221122 and 12231223, a subject might begin with the judgment that the strings are “fairly similar” since both can be split into identical halves: 1122+ 1122 and 1223 + 1223. This initial value might then be adjusted downward twice, once to account for the fact that the first string comprises two numerals while the second string comprises three, and once to account for the fact that the first string comprises pairs of identical numerals while the second does not.

It is important to note that the averaging model uses the information that is represented in the tree, but it does not use the formal properties of the tree itself. For example, in the judgment described above, the “salient” anchoring information is captured by the fact that both strings have repetition as their global (or last-applied) operator, and the two items of “adjustment” information are captured, respectively, by the facts that the strings differ in terms of the number of transposition operations that they involve, and in whether their local (or first-applied) operator is repetition or transposition. However, the topological structure or connectedness of the tree does not figure anywhere in the judgment process.

Tree-theoretic distance model

In the tree-theoretic model, a subject judges similarity by determining the length of the pathway between the nodes that are being compared: the shorter the pathway, the greater the similarity. Thus, the mechanism for judging similarity makes direct use of the structural characteristics of the representational scheme. (To the best of our knowledge, no formulation of the tree-theoretic model has specified exactly how the pathway is found and its length assessed. However, it is probably fair to suggest the sort of spreading-activation mechanism that has been hypothesized to occur for search in network- structures, e.g., Collins and Loftus 1975.)

The experiment

Method

Tasks

Three types of data were gathered in the experiment: (1) direct judgments of similarity among the eight types of string shown in fig. 1, (2) data from binary sorts of the eight string types, and (3) 5

protocols describing the bases for their sorts. The sorts and protocols were used in fitting the theoretical models to the similarity judgments.

The intent of the judgment task was that *Ss* judge the similarity of structure of the strings, and not simply item-by-item character identity. To facilitate this, *Ss* were asked to compare number strings with letter strings. Since this was a difficult task, *Ss* were given practice in describing and comparing the “patterns” of character strings before being asked to make direct judgments. (Data are presented below that confirm that *Ss* were able to perform the judgment task reliably.)

Stimuli and apparatus

The stimuli for the judgment task were the 64 possible pairings of the eight number strings and the eight letter strings. Of these, 56 were experimental pairs in which the two strings were not structurally identical, and 8 were anchor pairs in which the strings were identical except for the number/letter distinction. Each stimulus pair was typed onto a transparent sheet in horizontal format, i.e.,

11221122
abbcabbc

and mounted for presentation as a 35 mm slide. The number string always appeared above the letter string.

Slides were presented to *Ss* on a rear-projection screen in a sound-proof booth with presentation time being controlled by the *Ss* on a trial by trial basis. The response scale was an unmarked horizontal row of 80 dashes on a computer-controlled video monitor. The scale was anchored at the right (i.e. dash 80) with “very similar” and at the left (i.e. dash 1) with “not at all similar”. Above the line was a “response arrow” whose position could be controlled by a pair of buttons on a control console. A third button on the console transmitted the response to the computer and initiated the next trial.

The stimuli for the sorting task were two sets of eight 3X5 cards, on which the number strings and the letter strings were typed individually.

Procedure

Ss were run one at a time through two experimental sessions. Day 1 began with a pair of orientation tasks. In the first task *Ss* were shown four individual number or letter strings and asked to describe their patterns verbally. In the second task *Ss* were given three “multiple choice” problems for which they chose which of four letter strings was most similar, next most similar, and least similar to a number string, and then explained their choices to the experimenter. It should be noted that at no time did the experimenter attempt to “teach” the *Ss* any particular structure. The purpose of the orientation was solely to let *Ss* familiarize themselves with the idea that the stimulus strings could be described in terms of structure and compared with one another on that basis.

After the orientation tasks, *Ss* were taken to the experimental booth and shown how to use the response scale. They were asked to make direct ratings of similarity for three sets of stimulus pairs: one set of 44 pairs for practice followed by two full replications of the 64-pair stimulus set.

Interspersed between the three sets of stimulus pairs were two replications of the sorting task. For this task *Ss* were given a deck of either the eight number strings or the eight letter strings. These they were asked to sort into two piles [2] on the basis of similarity, with the sorts being recorded by the experimenter along with protocols of the *Ss*’ sort strategies. The two piles were then sorted further into smaller and smaller piles, and protocols were recorded for each successive sort, until a complete binary tree had been produced.

The procedure for Day 2 was identical except that the orientation tasks were omitted. This gave four replications per stimulus pair per 5 for the judgment task, and four replications of the sorting task—two with number strings and two with letter strings.

Subjects

The Ss were 15 undergraduate students from introductory psychology classes at the University of Wisconsin, Madison. The students received extra course credit for their participation.

Results

Performance on the judgment task

The experiment required that Ss be able to judge the structural similarity between number strings and letter strings. To the extent that they were able to do this reliably, their judgments should show no effect attributable to whether particular strings were presented as numbers or as letters. For example, if similarity is judged solely on the basis of structure, then the similarity between 11221122 and abbcabbc should be exactly the same as the similarity between 12231223 and aabbaabb since, for both these pairs, the structural comparison is the same.

To test this prediction, the data for each S were submitted individually to 2 X 28 analyses of variance in which the 28 unordered stimulus pairs constituted one factor, and the 2 levels of the letter/number format variable (see the example in the previous paragraph) constituted the other. The four within-cell replications gave an error term with 168 degrees of freedom. Of the 15 analyses of variance, none of the tests of a main effect for format and only one of the tests of an interaction between format and stimulus pair reached statistical significance. Since this is about what is expected by

Table 1

Subject	Intraclass correlations among sorts	RMSD distance model	RMSD averaging model	SEM raw data	F-ratio on difference ($df = 1, 56$)
1 ^a	0.67	9.34	7.68	5.26	7.28 ^b
2	1.00	8.63	5.68	6.83	8.93 ^b
3 ^a	0.80	5.96	4.29	4.81	7.07 ^b
4	0.82	10.86	11.50	6.63	1.35
5	1.00	9.70	8.18	6.82	2.05
6 ^a	0.96	7.03	7.51	5.09	0.77
7 ^a	0.84	7.55	4.84	4.05	14.86 ^b
8	1.00	7.86	8.72	7.93	0.71
9 ^a	0.80	12.13	5.30	7.88	37.16 ^b
10 ^a	0.97	3.51	3.67	4.42	0.56
11	1.00	9.50	6.93	6.27	6.10 ^b
12	1.00	9.16	9.33	7.34	1.40
13	0.95	7.92	5.39	7.15	5.91 ^b
14	0.99	12.12	9.13	10.56	4.60 ^b
15	0.98	11.41	7.12	8.04	11.62 ^b

^a Subjects using residual categories on at least one occasion during sorting.

^b $p < 0.05$.

chance, it is reasonable to conclude that Ss were able to judge the structural similarity between strings independently of the surface format in which the strings appeared.

Sorting task

In general, Ss appeared to be consistent in sorting the sets of character strings into trees. To get a quantitative measure of this, each S's trees were scored individually in terms of the number of nodes (or choice points) that separated the 28 possible pairs of strings. (To illustrate, in fig. 1 the string 11221122 is separated from the strings 12231223, 12121212, and 11112222 by one, three, and five nodes, respectively.) Then the intraclass correlation was computed across the four sets of scores. The results of the analyses are given in column 2 of table 1. Of the 15 correlations, five are 1.00 (indicating that these Ss sorted the strings identically on all four occasions), and the median value is 0.97. These high correlations reflect the fact that Ss tended to produce identical sorts from trial to trial, or to produce sorts that varied in only minor ways (that is, in ways that affected only a few of the 28 possible pairs).

Although the trees produced by individual Ss were very reliable over trials, there was marked variability between Ss in the structure of the trees they produced. For illustration, fig. 2 gives four trees

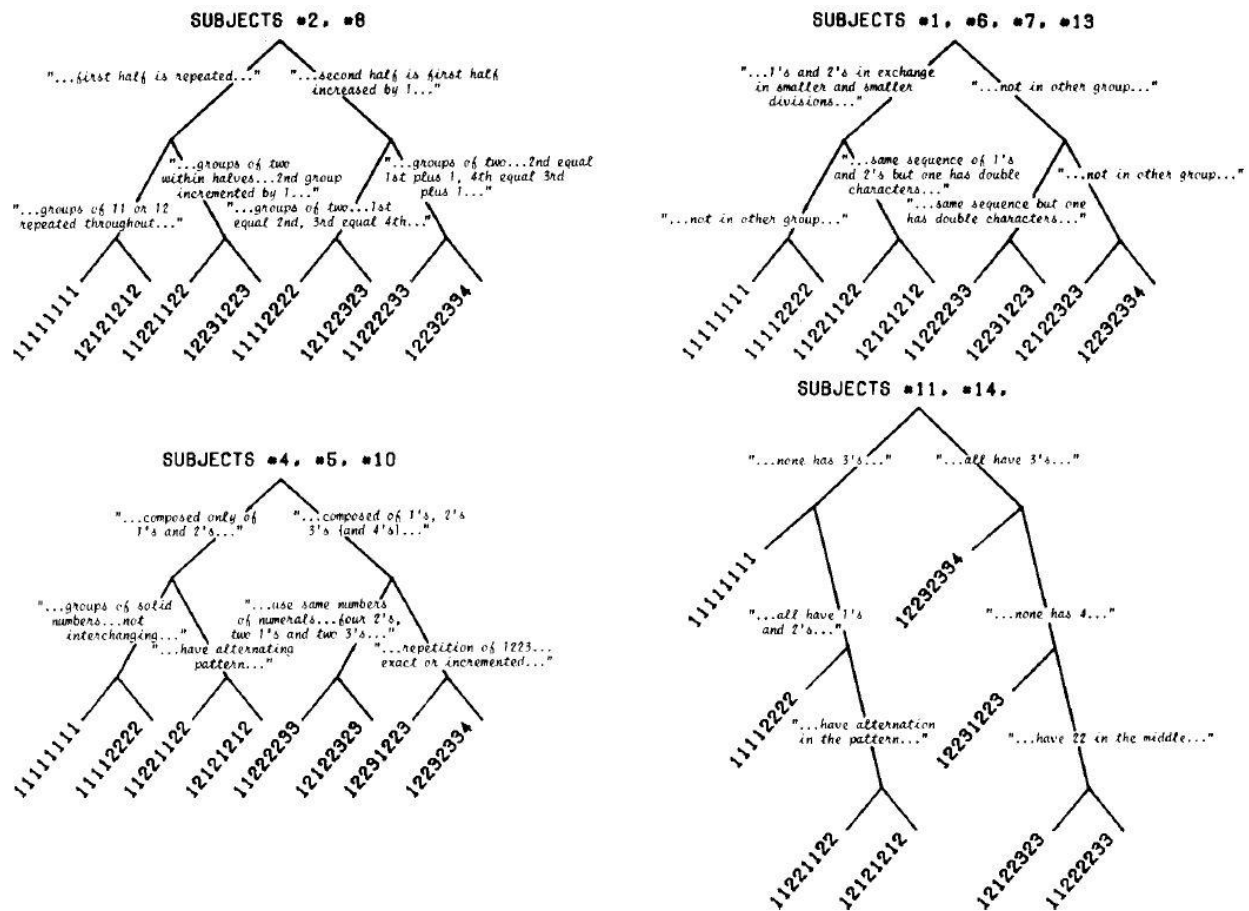


Fig. 2. Examples of sorting trees produced by subjects with samples of the comments used to describe the sorts.

that (taken together) are typical of the trees produced by 11 Ss, along with examples of the comments that the Ss gave in describing their sorting strategies. (The four Ss not shown in the figure each produced unique trees.) As can be seen, only Ss no. 2 and no. 8 (in the upper left corner) produced trees that were

identical to the stimulus tree given in fig. 1. Most other Ss (10 in all) produced trees that varied from the stimulus tree at even the most global level, that is, at the highest node of the tree. For these Ss, the first attribute that was used to distinguish among the strings was the number of different characters the strings comprised.

Since the numbers of Ss producing each of the various trees was small, it was not possible to determine whether there was any relationship between the tree structure an S preferred and the reliability of that preference over the four sorting trials. There was, however, a relationship between reliability in the sorting task and the types of sorting category that the S used (see footnote a in table 1). Ss who used only well-defined categories (that is, categories specified in terms of attributes that member-strings had in common) tended to be more reliable in the sorting task than Ss who also used ill-defined categories (that is, residual categories in which member-strings had nothing in common except not fitting into some particular well-defined category), $p = 0.06$ two-tailed by median test.

Fitting the models to the judgment data

For the tree-theoretic distance model, the data from the sorting task were used to find the best-fitting subjective stimulus tree for each S. (For 11 Ss, this was a simple matter since they had produced identical sorting trees on at least two of the four trials. For the remaining 4 Ss, the best-fitting tree was determined node-by-node by groupings the strings in accord with the groups most frequently produced by the Ss.) Once this had been done, the iterative curve fitting program STEPIT (Chandler 1969) was used with a least squares criterion to estimate the lengths of the 13 independent arcs [3] in the structural tree for each S, and predicted similarity values were obtained by adding the arc lengths in the paths connecting the various pairs of stimulus strings [4],

For the averaging model, verbal descriptions given by the Ss concerning the bases for their sorts were used to suggest stimulus dimensions that Ss might have used in the judgment task. Four dimensions figured in the descriptions of almost all Ss: three were directly related to the operations used in generating the strings, i.e., "the two halves are the same", "the strings show alternation", etc., and one reflected the degree to which the two strings used the same number of different characters, i.e., "all use only 1's, 2's, and 3's", etc.

Next, STEPIT (Chandler 1969) was used to fit the averaging model to the single S judgment data. Twelve free parameters were used for each S: four to describe the stimulus dimensions, seven to approximate weighting functions, and one to provide for left/right bias in the use of the response scale [5]. Since there had been no effect of the number/letter distinction on the judgment data, the similarity value predicted for a given pair of strings was the same regardless of which string was presented as numbers and which as letters.

Statistical comparisons of the models

Table 1 gives the single-S root mean squared deviations (RMSD's) between predicted and obtained judgments (relative to the 80 point response scale) for the tree-theoretic distance model (column 3) and the averaging model (column 4). For purposes of comparison, column 5 gives standard errors of the mean across the 56 stimulus pairs for each S. In general, both models were able to account for the judgment data fairly well.

Of the Ss, 10 were better fit by the averaging model and 5 were better fit by the tree-theoretic model. These results were evaluated statistically at both single S and group levels. Single S analyses were performed on the signed deviations between predicted and obtained judgments for the two models, using a 2X28 (model X stimulus pair) factorial design in which the format variable was used as a replications factor to get an error term. The results (shown in column 6 of table 1) indicated that for 9 of the 10 Ss who were better fit by the averaging model, the differences between models were statistically significant. In contrast, for the 5 Ss who better fit by the tree-theoretic model, the differences between models did not approach statistical significance.

To get a group test, analysis of variance was applied to the RMSD's in a 2X15 (model X subject) repeated measures design. The results of the test also favored the averaging model: $F(1, 15) = 11.26$, $p < 0.01$. Thus, with respect to goodness-of-fit, both single and group analyses indicated that the averaging model did a better job of accounting for the similarity judgments than the tree-theoretic distance model, and this despite the averaging model having used one fewer parameters.

Although the present results support the averaging model of similarity, it is important to emphasize that tests of goodness-of-fit such as those employed here are far from ideal and ought, where possible, to be replaced by tests of deviations from predictions (cf. Anderson and Shanteau 1977; Bimbaum 1973). The problem in the present case, however, is that the averaging model of equation (1) is completely nonlinear so that specific deviations-tests are not available.

But lack of better tests does not solve the problem that tests of goodness-of-fit are not unequivocal in demonstrating the superiority of one model over another. Thus, in the best case, they ought to be supplemented by other sorts of data. In a previous test of the averaging model (Lopes and Oden 1980) quantitative support for averaging drawn from tests of goodness-of-fit was supplemented by clear qualitative failures of the alternative model to account for the ordinal properties of the data. In principle, qualitative tests of this sort ought to be available also for the tree-theoretic distance model since ordinal predictions can be made about the ranking of similarity judgments for hierarchically organized stimuli. In the present case, examination of the raw data suggested many such violations of ordinal predictions. But these could not be tested due to the failure of our Ss to perceive a common hierarchical structure. Nevertheless it should be possible in future tests of the tree-theoretic distance model to get such ordinal data by using stimuli whose hierarchical structure is transparent and likely to be agreed upon by all Ss.

Discussion

Before proceeding to a discussion of some general issues bearing on representation and process in similarity judgment, a few words are in order concerning whether the finding that a model allows a good fit to data provides sufficient justification for giving psychological interpretations to the various parameter estimates produced by the curve-fitting program. Both practical and theoretical factors signal the need for caution in attempting such interpretations, as is particularly clear with respect to the present two models.

On the practical side are problems involved with noise in the data being fit and/or discrepancies between the hypothesized model of the data generating process and the "true" process that actually obtains in the world. Either factor can cause parameter estimates to be produced that are suspicious or unacceptable from a psychological point of view. In some cases precautions can be taken in fitting the model that prevent certain sorts of unacceptable parameters from occurring. For instance, distances in the tree model and weights in the averaging model were prohibited from taking on negative values since neither negative distances nor negative weights make sense psychologically. But some difficulties cannot be prevented. For example, in fitting the tree model to the data, arcs of zero distance were occasionally assigned for some subjects. Whether these zero distances are meaningful cannot be determined easily since they may have been caused by trying to fit noise or, as the goodness-of-fit results suggest, by using an inappropriate model.

The practical problems are even more serious for the averaging model. Although the unconstrained model is conceptually simple (containing only weights and scale differences), the tightly constrained approximation that was actually fit to the data (see footnote 5) has no ready psychological interpretation. While some of the particular approximations used may be psychologically reasonable (i.e., the assumption that weights vary as a simple function of scale difference), some of the actual parameters manipulated by the curve-fitting program (e.g., the coefficients in the polynomial used to

approximate the weighting function) have no psychological content whatsoever. Thus, the best one can say is that the model fits reasonably well in spite of necessary constraints, and that loosening such constraints by using more parameters could only improve the goodness-of-fit.

On the theoretical side are problems concerned with finding a vehicle for presenting or displaying parameter estimates that is both theoretically appropriate and intuitively compelling. For example, arc lengths in the distance model could be displayed graphically in tree-diagrams. This would be simple, effective, and theoretically suitable for capturing the role played in the model by the parameter estimates. But things are not easy for the averaging model. In the simplest treatment, one might assume that both scale differences and weights are real subjective quantities that are intrinsically associated with particular stimulus pairs, this would allow weights and scale differences to be construed as points on underlying subjective dimensions. However, such dimensions would bear no particular structural relationship to one another and so would not lend themselves to spatial or pictorial representation. For the anchoring-and-adjustment conception of averaging, things are even worse. Since in that view weights reflect in part the order in which operations are performed on scale differences, they are only defined in the ongoing context of processing particular stimulus configurations. Thus, a proper interpretation of weight parameters would need to convey a sense of the dynamic properties of the judgment process. This is by no means a simple matter, although eye movement recordings might be useful (Russo and Rosen 1975).

Cognitive representation of structural strings

The averaging model of similarity judgment assumes that the process of *judging* similarity is not merely a byproduct of the structure in which information is represented in the mind. But this is not to say that structures for representing knowledge are uninteresting or unimportant. It is quite clear that human beings know a great deal about the relationships that can exist among and between various objects and concepts.

In the present experiment we operated under the tacit assumption that subjects would perceive the stimulus strings to be organized hierarchically. This assumption is worth examining. The tree shown in fig. 1 represents the design of the stimulus set we employed. The particular order of the levels, from global at the top to local at the bottom, was chosen because it is useful for describing how the strings were generated. Nevertheless, this particular tree is only one of the structures in which repetition and transposition operators can be organized to yield the eight stimulus strings. Alternative trees can be produced using any of the possible orderings of the three levels, and, as it turns out, it is even possible to obtain a non-hierarchical interpretation of the stimulus set by assuming that the strings lie at the eight corners of a cube in which the three spatial dimensions (each having repetition at one end and transposition at the other) corresponds to the three levels of the tree;.

Two features of the sorting data argue that subjects did perceive the stimulus strings to be organized hierarchically. First, all but two of the subjects produced sorting trees in which the tests differed from one node to another. This is important since only trees that have identical tests at each choice point can be mapped onto coordinate spaces that have no more dimensions than the tree has levels [6]. Second, subjects overwhelmingly and reliably preferred sorting strategies that were consistent with the sorts of hierarchical organization that appear to be involved in perceptual representation (Palmer 1977). In particular, subjects typically produced trees in which global string characteristics (such as the number of characters in the complete string or the presence of a pattern involving the two halves of the string) dominated local string characteristics (such as a transposition of adjacent characters).

Processing structure

In research using structural strings of the sort employed in the present experiment, Restle (1970) found that errors made by subjects when learning to reproduce long binary patterns were related in predictable ways to the structural organization of the stimulus trees used to generate them originally. This prompted us to assume incorrectly that subjects would tend to use exactly the same sorts of structural information in judging the similarity between pairs of such strings. Obviously, the data proved otherwise. In retrospect, it is not surprising that subjects use different structural information in serial learning than they do in similarity judgment. In fact, the single most important point this paper can make is that psychological processes like learning or similarity judgment are logically separate from the data on which they operate, so that structural factors that may be of major importance for one process may have little or no relevance for another.

The averaging model of similarity judgment differs fundamentally from most alternative models for similarity, including the tree-theoretic model treated here and the multidimensional distance model treated elsewhere (Lopes and Oden 1980), in the relative importance it attaches to the structures, on the one hand, that people use to represent what they know about the world, and the processes, on the other hand, that they use to operate on these represented data. Taken by themselves, the sorting data suggest that hierarchical trees provide a useful framework for describing what subjects know about the structural properties of the character strings used in the present experiment. But the fact that the averaging model did a better job than the distance model of fitting the judgment data suggests the possibility, at least, that the tree structures themselves may not play any functional role in the process of judging the similarity between pairs of such strings.

References

- Anderson, N.H., 1974. 'Information integration theory: a brief survey'. In: D.H. Krantz, R.C. Atkinson. R.D. Luce and P. Suppes (eds.). *Contemporary developments in mathematical psychology*. Vol. 2. San Francisco, CA: W.H. Freeman.
- Anderson, N.H. and J.C. Shanteau. 1977. Weak inference with linear models. *Psychological Bulletin* 84, 1155-1170.
- Birnbaum. M.H., 1973. The devil rides again: correlation as an index of fit. *Psychological Bulletin* 79, 239-242.
- Chandler, J.P., 1969. Subroutine STEPIT—finds local minima of a smooth function of several parameters. *Behavioral Science* 14, 81-92.
- Collins. A.M. and E.F. Loftus, 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82, 407-428.
- Cunningham, J.P., 1978. Free trees and bidirectional trees as representations of psychological distance. *Journal of Mathematical Psychology* 17. 165-188.
- Cunningham, J.P. and R.N. Shepard, 1974. Monotone mapping of similarities into a general metric space. *Journal of Mathematical Psychology* 11, 335-363.
- Johnson, S.C., 1967. Hierarchical clustering schemes. *Psychometrika* 32, 241-254.
- Kruskal, J.B., 1964a. Multidimensional scaling by optimising goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1-28.
- Kruskal, J.B., 1964b. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29. 115-129.
- Lopes, L.L. and G.C. Oden, 1980. Comparison of two models of similarity judgment. *Acta Psychologica* 46, 205-234.
- Palmer, S.E., 1977. Hierarchical structure in perceptual representation. *Cognitive Psychology* 9, 441-474.
- Restle, F., 1970. Theory of serial pattern learning. *Psychological Review* 77. 481-495.

- Russo, J.E. and L.D. Rosen, 1975. An eye fixation analysis of multialternative choice. *Memory and Cognition* 3, 267-276.
- Shepard, R.N., 1962a. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* 27, 125-139.
- Shepard, R.N., 1962b. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* 27, 219-246.
- Shepard, R.N. and P. Arabie, 1979. Additive clustering: representation of similarities as combinations of discrete overlapping properties. *Psychological Review* 86, 87-123.
- Tversky, A., 1977. Features of similarity. *Psychological Review* 84, 327-352.
- Tversky, A. and D. Kahneman. 1974. Judgement under uncertainty: heuristics and biases. *Science* 185. 1124-1131.