DOCUMENT RESUME

ED 079 800 CS 500 367

AUTHOR Simpkins, John D.

TITLE Recoding Numerics to Geometrics for Complex Discrimination Tasks; A Feasibility Study of Coding Strategy.

SPONS AGENCY National Center for Health Services Research and Development (DHEW/PHS), Rockville, Md.

PUB DATE Apr 73

NOTE 16p.; Paper presented at the Annual Meeting of the International Communication Assn. (Montreal, April 25-29, 1973)

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Cybernetics; *Display Systems; *Engineering Technology; Information Processing; Information Retrieval; Information Science; Information Systems; *Information Theory; Information Utilization; *Medical Treatment; Methodology; Models; Pattern Recognition

ABSTRACT

Processing complex multivariate information effectively when relational properties of information sub-groups are ambiguous is difficult for man and man-machine systems. However, the information processing task is made easier through code study, cybernetic planning, and accurate display mechanisms. An exploratory laboratory study designed for the University of Missouri School of Medicine and the Department of Electrical Engineering, after pretesting by the Department of Information Science, sought feasible coding strategies for displaying multivariate biochemical data gathered on a set of patients. The coding strategies served as "facilitators" of the human perceptual process to permit an accurage placement of the patients into two groups--normal and diseased. Geometrical designs successfully functioned as "codes" for a variety of body chemistry states. Users were able to interpret the visually displayed designs because of a natural human facility for pattern discrimination and recognition. The results of this laboratory test encourage further work of this kind. (CH)

Recoding Numerics to Geometrics for Complex Discrimination Tasks

A Feasibility Study of Coding Strategy

John D. Simpkins

This manuscript has been prepared for the Information Systems Division, International Communication Association, Annual Meeting, April 25-28, 1973, Montreal, Canada.

Recoding Numerics to Geometrics for Complex Discrimination Tasks

A Feasibility Study of Coding Strategy*

John D. Simpkins**

Processing complex multivariate information accurately and effectively when relational rules (or properties) of information sub-groups are ambiguous or unknown is indeed a difficult task for both man and man-machine systems.

Perhaps such an information processing task can somehow be influenced by the manner in which the information is received at both the sensory and perceptual levels of the human observer. The selection of the code for the message may carry some major responsibility for achieving communication tasks. Bruner,[1] Kidd,[2] and others have presented both data and suggestions in support of such a generalization.

The implications of the generalization rather than the principle itself are of greater interest here. Namely, that as the observer's or perceiver's tasks or goals change, it may be desirable to alter the nature of the coding strategy to "best fit" the goals as they occur (e.g.) serially or simultaneously. Also, as information processing in man-machine systems progress from input to subsequent understanding of output, it may be useful to select alternative modes of displaying information.

---

*The research reported here was supported by USPHS Grant HS 00014 from the National Center for Health Service Research and Development.

**Dr. Simpkins, currently Assistant Professor, College of Communication Arts, Michigan State University, East Lansing, was a Research Associate, Department of Information Science, University of Missouri-Columbia, when this research was completed.

[1]Jerome Bruner _et al._, _A Study of Thinking_, Wiley, 1956.

[2]J.S. Kidd, "Human Tasks and Equipment Design," in _Psychological Principles of Systems Development_, R.M. Gagne (ed.), Holt, Rinehart Winston, 1966, p. 177.

In situations such as the latter, it's likely that at each stage in the processing there are different human goals and different "user" groups which bring various skills and abilities to the communication situation.

Criteria for Coding Strategy

Kidd reported in 1966 "an increasing tendency to employ computer display combinations" in multivariate decision tasks.[3] There is little indication of any change in that trend. The application of graphic displays to problems of medical diagnostic classification is central to the feasibility study reported below. However, before proceeding to the study, it seems appropriate to identify two major principles that serve as guides in the selection of coding modes and strategies in communication, and information processing, in particular:

1. that the information be presented in a form appropriate to the communication skills of the recipient; and

2. that the information be represented in a form appropriate to and compatible with the objectives or goals of the communication activity.

Harrison[4] and others have discussed the importance of the role of non-verbal activity and communication in human interaction. In most instances, such encoding is done simultaneously with verbal encoding and has been referred to as "metacommunication." Bateson has termed such encoding as extraverbal activity.[5] Both Harrison and Bateson have discussed how important it is to provide support of verbal communication efforts with other message and coding

---

[3] Ibid.

[4] Randall Harrison, An Introduction to Non-Verbal Communication, Prentice-Hall, in press.

[5] Gregory Bateson, "Information, Codification and Metacommunication" in Communication and Culture, Alfred G. Smith (ed.), Holt, Rinehart Winston, 1966.

strategies (e.g. gestures, facial expression, silence, etc.) to improve the likelihood that communication goals will be achieved. Recoding information to achieve communication or to increase the probability of its occurrence in human interaction is not an uncommon practice.

Such recoding may or may not be complete transformation. George Miller has suggested that "mnemonic devices (partial transformations) are frequently used . . . for increasing the amount of information that we can deal with."[6] Recoding in this sense is a re-organization of basic communicative elements facilitated by what Miller has described as "chunking." However, recoding of this type does not necessarily require a change of the nature of the code and its elements, but suggests itself more as an alteration in information handling strategy by the recipient.

The study reported below examines the feasibility of recoding as a symbol substitution process to facilitate complex classification tasks on the basis of relational properties of data. Numeric data have been recoded as geometric configurations (see Plate 2) on the assumption that such patterns may make more visible the salient classification properties of the data. These properties, it was reasoned, may not be as visible to the final user when the information is seen as numbers, especially when the numbers occur in combinations of four or eight measured variables, all on largely different scales.

Measurement models used for quantative coding operations carry with them certain assumptions and relational properties. Mapping numbers to events is primarily a matter of maintaining the relational properties (or the information) of the events measured. Once this coding transformation is completed, it provides certain major advantages for data control, manipulation

---

[6]George Miller, Psychology of Communication, Penguin Books, p. 40

and analysis. As such, the coding process utilized to "map" the events is appropriate for goals of analysis. However, the question of interpretation, application, and understanding of the analysis may be another matter, especially among recipients unfamiliar with or unaccustomed to the numeric codes, or their underlying models.

Feasibility Study

An exploratory laboratory study designed for S groups at the University of Missouri-School of Medicine and Department of Electrical Engineering were first pre-tested with Department of Information Science Staff. The purpose of the study was to search for feasible coding strategies for displaying multivariate biochemical data gathered on a set of 34 patients from the University of Missouri Medical Center in Columbia.

Coding strategies were to serve as "facilitators" in the human perceptual process to permit an accurate placement of the patients into two groups --- one normal and one diseased. Also, the coding strategies were to focus on presentation of the data such that the observer could create a "gestalt" of the patient as reflected through the measured biochemical variables. In other words, the investigators were concerned with somehow making the "structure" of the data base visible to the observer and in a form to capitalize on the high level capability of the human information processing system for pattern discrimination.*

*Other investigators on the project are experimenting with varieties of math models that might be fitted to the data base to separate and classify the 34 patients.

The study was designed in two phases. However, only the Phase II study is described in detail below.

Materials. Materials used were of two types: cathode ray tube pattern displays on an IBM 2250 similar to Plate 1. (for Phase I); and two sets of 34 3x5 index cards -- one set with two geometric figures on each card, and one set with four figures (for Phase II). Plates 2 and 3 illustrate cards from each of the two sets.

Phase I. Forty sets of three geometric patterns comprised the exercise displayed on the 2250. The purpose of this phase of the study was simply to examine Ss preferences for certain geometric dimensions that were used for coding information in the Phase II task. In effect, this phase was simply to explore possible control variables for later studies. While Plate 1 provides an example of the materials used in the (Phase I) preference study, there are two differences between the plate and the actual CRT display: the actual displays were not enclosed in a rectangular box, and the geometric figures were constructed of dot patterns instead of solid lines.

Insert Plate 1 about here

Phase II. The stimuli used in the card sorting task are illustrated by Plates 2 and 3. Each card represents measures of one of the 34 patients. The size of each of the geometric figures represents the magnitude of the variable assigned to each of the geometric forms and the lateral position of each of those figures on the card (from left to right) represents the magnitude of the remaining variables.

Insert Plates 2 & 3 about here

Therefore the card with two geometric forms displayed represents patient data on four measured variables and the card with four geometric forms representspatient data on eight measured variables. The assignment of the variables was as follows:

Size:	square = potassium	Position:	square = phosphorous
	triangle = sodium		triangle = calcium

A comparison of patients 1 and 2 as displayed in Plate 2 illustrates the potassium values for the two patients are similar, and sodium values are also similar. However, phosphorous is less for patient 1 than for patient 2, while patient 1's calcium level is greater than patient 2's.

Assignment of the eight variables was as follows:

Size:	square = potassium	Position:	square = phosphorous
	triangle = sodium		triangle = calcium
	hexagon = chloride		hexagon = BUN
	circle = bicarbonate		circle = creatinine

Plate 3 illustrates that the potassium values of the two patients is approximately the same, as are their sodium values. Chloride is a much lower value for patient 1 than for 2, while bicarbonate is greater and phosphorous is less for patient 1. Calcium is greater for patient 1 while BUN and creatinine levels are both less for patient 1 than for 2.

It is important at this point to make clear that the original numeric values of the eight variables were not on scales with the same minimum and maximum values. Therefore, the values were transformed to standardized scales before the form sizes and positions were determined for each of the variables. It is also important to note here that the recoding strategy did not reduce measured equal interval data to ordinal data even though the actual differences in the sizes and lateral positions of the forms are now a matter of perceptual judgment rather than a set of arithmetic operations.

Procedures. All Ss were simply instructed to sort the 2 sets of 34 cards into two piles such that the cards in each pile were "Alike" or "went together". Ss were not advised at any time what the cards represented, nor were they given any decision rule to apply to the sorting task. A record was kept for each S indicating the number of cards grouped together as well as which specific cards comprised each group for each of the 4 and the 8 variable decks.

Also, each of the two S occupational groups were divided so that one-half of each group sorted the 4 variable deck prior to the 8, and vice versa for the other half of the Ss.

Subjects. Thirty-one medical school students at the University of Missouri-Columbia participated in this research. Twenty graduate students in the Department of Electrical Engineering volunteered for the study. Total Ss = 51.

Data Analysis. The data collected for each S for each of the card sort tasks were "correct" and "incorrect" classifications of patients as "diseased" or "normal". Because Ss were not advised of the nature of the two groups, but were instead instructed to simply sort each deck into two groups, the number of correct classifications was determined by E assigning the label "diseased" or "normal" to that card stack which contained the greatest number of each of those patients. This procedure made the scoring such that 50% correct was the poorest obtainable performance. The data from the card sorts were tabled as illustrated in Figure 1.

Insert Figure 1 about here

Fisher Exact Probability Tests were computed for each matrix (or each S) to determine the number of significant matrices. After computing the FEP tests, a "score" for each S was calculated using the log of the cross product ratio of the 2 x 2, a procedure that would in part solve the problem of unequal marginals resulting from the Ss'performance. (S's were unaware that of the 34 patients, 17 were normal; 17 were diseased.) The log score was used as relative index of performance, i.e. the higher the number the greater the number of patients correctly classified. (Minimum value = 0, which corresponds to the 50% minimum correct; Maximum value = to 7.1, for 100% correct classification.) A test of difference between the occupational group was accomplished with a t test of classification scores.

Data Analysis. The analysis was primarily exploratory. A test of the occupational variable failed to distinguish the two groups on the classification card sort task. Therefore, the two groups were pooled to facilitate a better understanding of the influence of the nature of classification practise on subsequent classification tasks. As was true in the pre-test some Ss performed the sorting task with the 4 variable deck prior to the 8 variable deck, and some the reverse. (Recall, too, that the 8 variable deck contained the same 4 variables that appeared along in the 4 variable deck.)

General findings. Overall, practise with the eight variable deck enhanced S performance with the four variable deck, but the reverse was not true. Also, the number of significant (FEP Tests) tests of Ss classification task was greater with the eight variable than the four variable exercise. Such a finding is compatible with findings reported earlier by Bieri.[7] There were, among the total of 51 Ss, 9 significant matrices for the 4 variable deck and 30 for the 8

[7] Bieri, J. et al. Clincial and Social Judgment, Wiley and Sons, 1966, p.64.

variable deck. Additionally, it should be noted that all of the significant matrices created by the 4 variable deck exercises were produced by Ss who did equally well with the 8 variable deck, under both order (sorting task sequence) conditions.

Because of the feasibility nature of the study, specific details of the study results are of less interest than the more general observations to be made. There are no specific findings reported in the complete study document[8] to discourage the investigator from continuing a more systematic inquiry into the application of recoding strategies such as those employed here.

Some general findings support this enthusiasm: in general, Ss are able to successfully make these complex discriminations; more Ss performed the classification task more effectively with the 8 variable deck (the high information field) than with the 4 variable deck. (However, there may be some statistical reasons for this performance.) The success of some of the Ss is sufficient to suggest that it is indeed feasible to use such a coding strategy or some variant of the strategy, for displaying complex multivariate data for classification purposes, even when Ss are unaware of the content of the patterns, and the nature of the relations among the variables in ambigious or unknown. Performance among the Ss ranged from slightly more than 50% to a maximum of 92% correct classifications of the 8 variable deck. Three Ss were able to correctly classify the 34 cards with greater than the 90% correct level. The extent of the variability is encouraging and the ease with which such tasks are performed is also encouraging.

It's a matter of speculation at this point about the relative

---

[8] J.D. Simpkins, and D. Lindberg, Disease Classification as an Exercise in Human Pattern Recognition, Preliminary Report, Information Science Series, Documentation Note, University of Missouri-Columbia, November, 1972, Columbia, Mo.

classification success that might have occurred for medical students and engineers had defining content information been available to them. Subsequent studies with medical professionals might examine such problems using awareness as an independent variable to estimate the impact of variables such as knowledge, theoretic training, and others on the classification of cards such as those used in this study to represent patients recoded from numerical data to geometric configurations. Further studies will also provide for comparative appraisals of this coding method.

Such a code can be easily and readily used on CRT's in combination with more conventional numeric displays if later research suggests some defining criteria for such usage. It is, too, in part, the ease with which this coding strategy can be adapted to CRT usage that encourages future research.

BIBLIOGRAPHY

1. Bateson, Gregory, "Information, Codification and Metacommunication", Communication and Culture, New York, Holt, Rinehart & Winston, 1966

2. Bieri, J. et al., Clinical and Social Judgment, New York, Wiley & Sons, 1966

3. Bruner, Jerome et al., A Study of Thinking, N.Y., Wiley and Sons, 1956

4. Harrison, Randall, An Introduction to Non-Verbal Communication, Englewood Cliffs, N. J., Prentice-Hall, in press

5. Kidd, J.S., "Human Tasks and Equipment Design," in Psychological Principles of Systems Development, New York, Holt, Rinehart & Winston, 1966

6. Miller, George, Psychology of Communication, Baltimore, Md., Penquin Books, 1967

7. Simpkins, J.D., and D. Lindberg, Disease Classification as an Exercise in Human Pattern Recognition, Preliminary Report, Information Science Series, Documentation Note, University of Missouri-Columbia, November, 1972, Columbia, Mo.

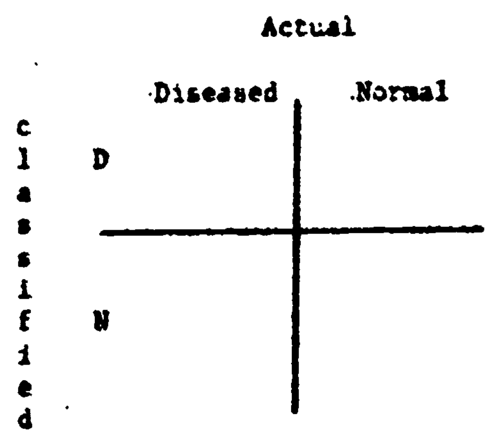

Figure 1

Analysis Matrix
for Card Sort Task

Plate 1

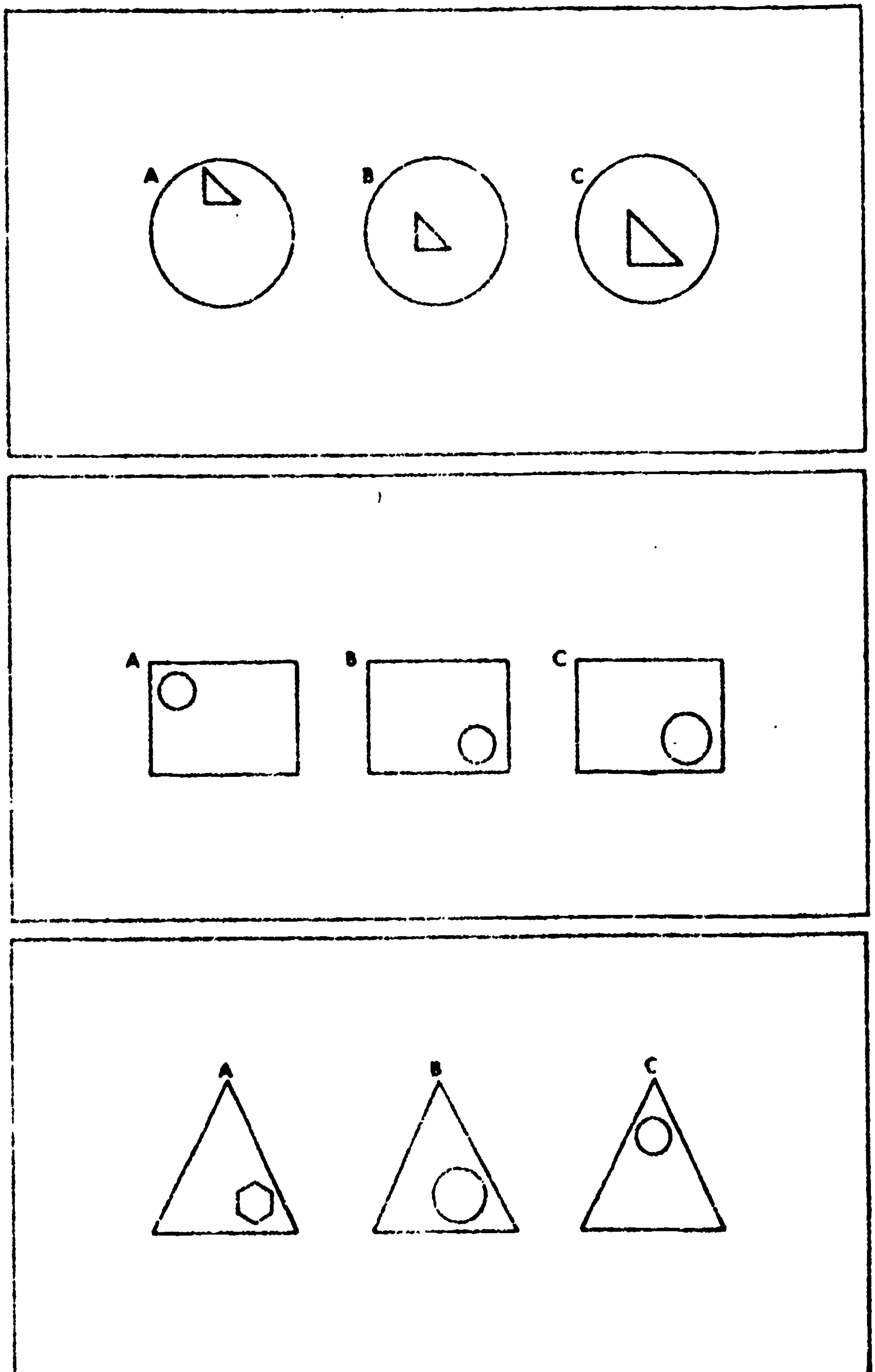

Plate 2

Plate 3