TRUTH TABLE CLASSIFICATION AND IDENTIFICATION*

EUGENE W. RYPKA

Department of Microbiology, Lovelace Foundation for Medical Education and Research, Albuquerque, N.M. 87108, U.S.A.

(Received 15 July, 1971)

Abstract. A logical basis for classification is that elements grouped together and higher categories of elements should have a high degree of similarity with the provision that all groups and categories be disjoint to some degree. A methodology has been developed for constructing *classifications* automatically that gives nearly instantaneous correlations of character patterns of organisms with time and clusters with apparent similarity. This means that automatic numerical *identification* will always construct schemes from which disjoint answers can be obtained if test sensitivities for characters are correct. Unidentified organisms are recycled through continuous classification with reconstruction of identification schemes. This process is cyclic and self-correcting. The method also accumulates and analyzes data which updates and presents a more accurate biological picture.

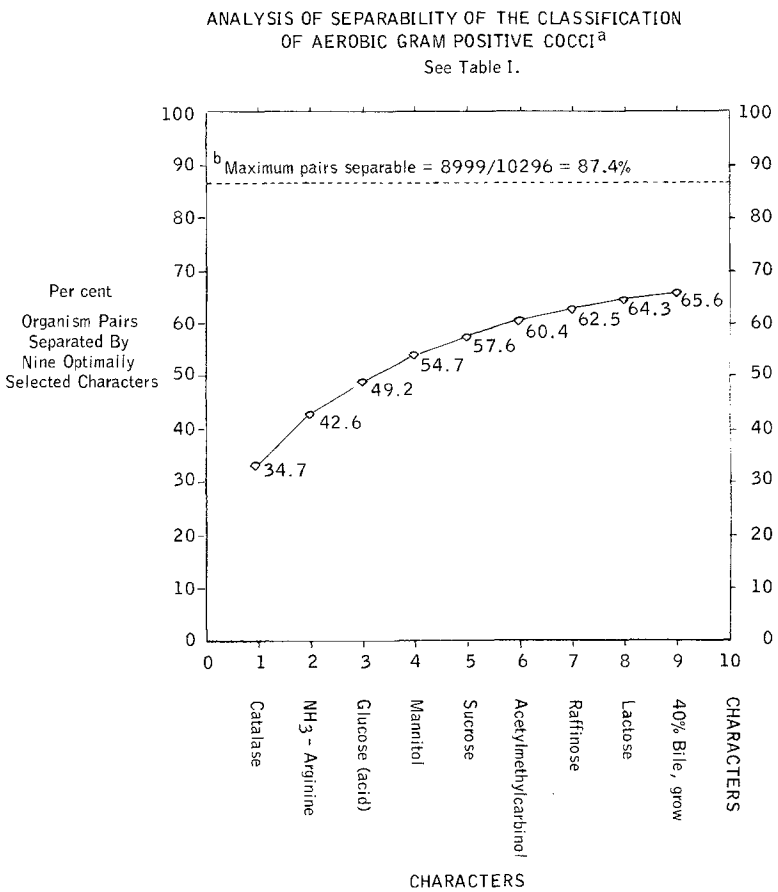
1. Introduction

There is an innate human ability and need to classify elements (and events) and to identify future elements as belonging to some class, group, set, cluster, or category which has been made previously. If identification of an element is impossible, formerly grouped elements may be reclassified to accomodate the unidentifiable element(s). This indicates that classification was initially *imperfect* (Jevons, 1877) or *artificial* – meaning all possible elements were not available for consideration. *Perfect* (Jevons, 1877) or *natural* classification would be possible if all elements were available for study at one time. The ability of the mind to group or to classify elements into clusters on the basis of 'apparent similarity' (Caws, 1965) is recognized but not totally explicable. This ability means some readily evident attributes, traits, or characteristics of the elements are recognizable as being commonly present among members of the cluster or group. The fewer the properties, attributes, traits, or characters of elements which are considered, the higher the 'abstraction level' of the group (Hayakawa, 1964). The greater number of properties, attributes, traits, or characters of the elements considered, the lower the 'abstraction level' of the group. The abstraction of the group may be represented by a symbol, a common noun, or a naming word. Psychologists may consider classification in terms of concept learning (Hunt, 1962), logicians and mathematicians in terms of symbolic logic (Styazhkin, 1969) and set theory (Zermelo, 1908), and biologists in terms of evolution.

Later consideration will be given to the problems of defining the functions of classification and identification. For the present – and to approach classification and identification at a simple, practical, operational level – they will be considered inverse

* This work was supported by Grant AI 16385–02 and 16385–03 National Institutes of General Medical Sciences.

Space Life Sciences 3 (1971) 135–156.



[a] Matrix dimensions: 144 organisms x 157 characters.

[b] Total organism pairs to separate: $G(G-1)/2 = 144(144-1)/2 = 10296$.

[b] Maximum pairs separable by 91 characters: 8999/10296.

Fig. 1. In an organism versus character matrix (Table V), if the character values of every possible pair of organisms are compared, those characters that separate the pair may be found (Figures 2 and 3). Also, those pairs of organisms that are not logically disjoint (meaning at least one character present for one member and absent for the other (see also Lapage, 1970)) may be found (Table I). In this figure the rate of separation of pairs of organisms by successive use of nine optimally selected characters is illustrated. Note of 10296 possible pairs, only 8999 pairs (87.4%) are separable. This indicates – *for the characters used* – the classification is not logically disjoint. This represents a reasonably exhaustive analyses. See Table I.

mental activities of imperfect induction and of deduction (Jevons, 1877), respectively. In this discussion, a method will be illustrated of forming clusters of elements with a high degree of similarity, with the provision that each cluster (and its elements) will be disjoint or separable to some degree from other clusters (and their elements) (Rypka, 1971a, b). Identification schemes constructed (Rypka *et al.*, 1967) from such classifications will always assure that every element is disjoint, provided test sensitivity for determining the presence or absence of character(s) is adequate. That classifications are constructed without logically disjoint elements is demonstrated in Figure 1 and in Table I using illustrations from microbiology (Rypka and Babb, 1970b, c).

The method of classification to be described also creates optimal total linkage of character patterns of elements with time in groups or clusters with evident or apparent similarity (Rypka, 1971a). The method may be used without interruption or intermittantly and is called continuous or truth table classification. It must be emphasized that the method to be described creates a practical, operational type of classification which may be utilitarian on an interim or other bases. No other claims are made for this invention and logically all methods go back to antiquity.

2. Methods

Propositions. Propositional functions become propositions when values are assigned to the variables. Thus, propositional functions of the form

Propositional functions			Propositions		Probability
Subject	Predicate		Subject	Predicate	
All	x is	y	All bacteria	are motile	$p = 1$
No	x is	y	No bacteria	are motile	$p = 0$
Some	x is	y	Some bacteria	are motile	$0 < p < 1$
Some	x is not	y	Some bacteria	are not motile	$0 < p < 1$

become propositions when values are assigned to x and to y (Hilbert and Ackerman, 1950). The method of continuous or truth table classification demonstrates how *propositional predicates* representing organism characters in matrices are rearranged horizontally in descending order depending upon the number of characters present over all organisms (Table VI), and how the *propositional subjects* representing elements or organisms are rearranged vertically in ascending order of the truth table number (Table VIII).

The propositional functions 'some x is y' and 'some x is not y' are used to designate variable characters. Variable characters should not exist for single elements, that is, one element's character values should be 1's and/or 0's (see Table II for symbols and their meanings) if uniform testing for characters have been done and interpretations of test results are not ambiguous. However, when logical summaries (Table III) are made of the character values of elements in clusters, variable character results are possible.

TABLE I

An analysis of 908 heterotrophic bacteria and 724 of their characters. The organisms are placed in eight subgroups depending upon their oxygen requirement, Gram staining characteristic, and cell morphology. Depending upon character variation in one or more of these three characters an organism could be placed in all eight subgroups. The per cent disjointness or separability of each subgroup is shown at right. The analysis is reasonably exhaustive (Rypka and Babb, 1970b, c).

Subgroups[a]	Original Matrix Dimensions[b]	Pairs Separated By Characters Used	Pairs To Separate[c]	Per cent Pairs Separated
1. Anaerobic Gram negative cocci	8x91	13/	28	46.429
2. Anaerobic Gram negative bacilli	80x118	2256/	3160	71.424
3. Anaerobic Gram positive cocci	42x88	506/	861	58.769
4. Anaerobic Gram positive bacilli	181x166	14433/	16290	88.613
5. Aerobic Gram negative cocci	82x95	2746/	3321	82.716
6. Aerobic Gram negative bacilli	244x252	26296/	29646	88.653
7. Aerobic Gram positive cocci	144x157	8999/	10296	87.354
8. Aerobic Gram positive bacilli	177x166	12757/	15576	81.947

[a] Analysis covers 908 heterotrophic organisms and 724 characters. An organism may occur in more than one subgroup.

[b] Matrix dimensions refer to the number of organisms by the number of characters. Only characters which, for at least one organism in the matrix, have a character definitely described as being absent, variable, or present are included.

[c] Pairs to separate = G(G-1)/2 (G = no. of organisms in the subgroup).

TABLE II

The symbols and their meaning used in the text and in Tables II through XVII.

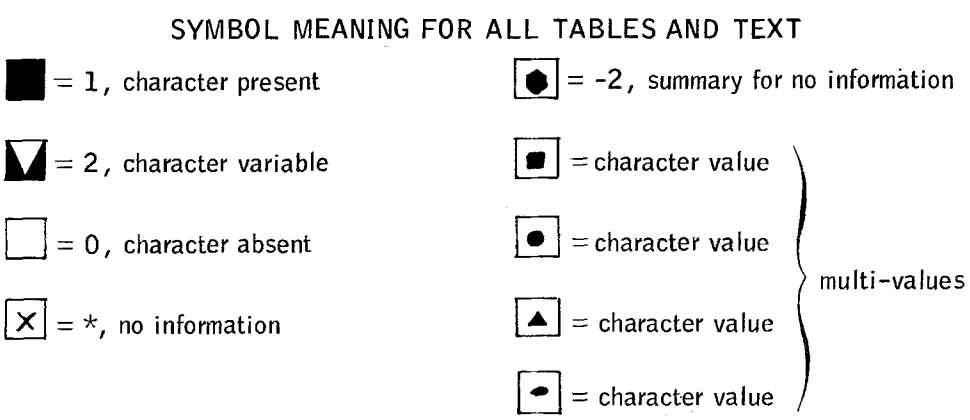


TABLE III

The method of preparing summaries of character values from literature (Rypka and Babb, 1970a).

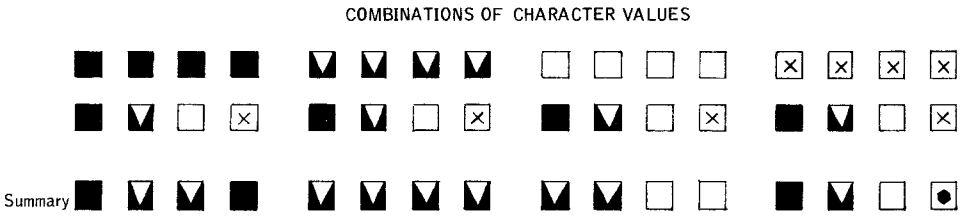


TABLE IV

The left matrix illustrates a truth table constructed for four two-valued characters. Every possible combination of values is shown and all binary sequences are different. The right matrices illustrate three four-valued characters. Each symbol may represent a range of quantitative values, for example, the per cent utilization of different substrates/bacterial unit/time.

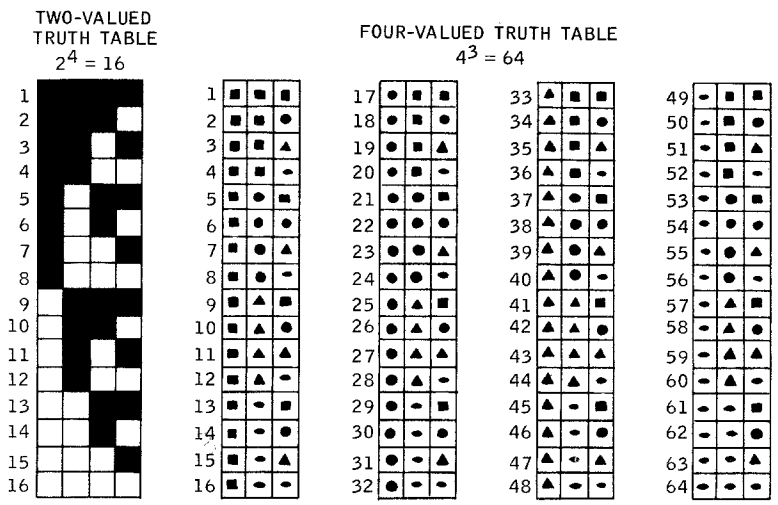


TABLE V

A random organism versus character matrix showing the order in which data were collected about organisms. Dimensions: 20 organisms by 8 characters. The number of times each character is present, considered over all organisms in the matrix, is n_1, and is shown at the bottom of the table. The characters are reordered in descending order of n_1 values in Table VI (Rypka, 1971a).

RANDOM MATRIX

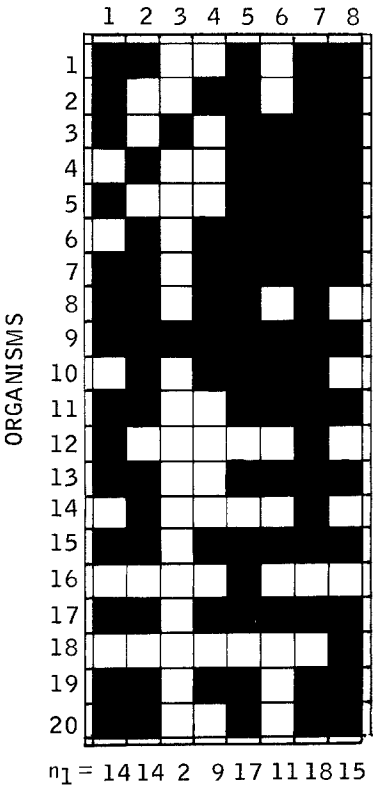
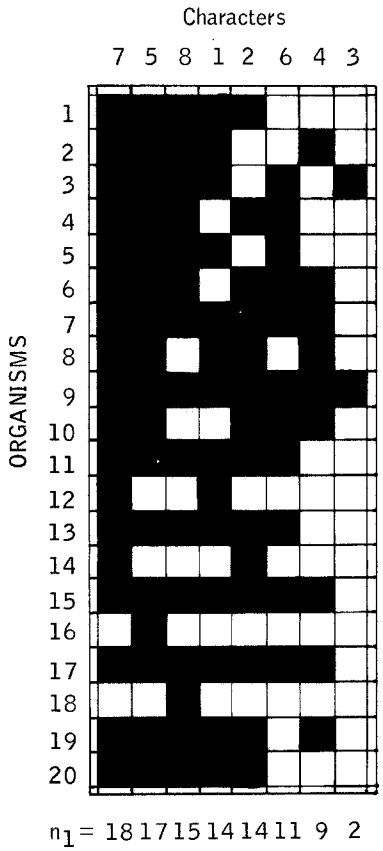


TABLE VI

The horizontal non-random matrix created by rearranging the characters in Table V in descending order of their n_1 values. The character values, for each organism and in the order shown, are traced on a truth table with the same number of two-valued characters (eight in this example and $2^8 = 256$). Each organism's character values will fit one and only one of the 256 patterns. See Table VII and Table VIII (Rypka, 1971a).

HORIZONTAL NON-RANDOM MATRIX



n_1 = number of characters present over all of the organisms

Thus, two-valued logic (character present = 1, absent = 0) or three-valued logic (character present = 1, variable = 2, absent = 0) may be used in truth table classification. It is emphasized that many-valued logic (Lukasiewicz, 1920; Rosser and Turquette, 1952) is possible (Table IV); this will be briefly discussed below in this section under *Truth tables.*

Matrices. A *random matrix* (Table V) is an unordered table of elements or organisms (propositional subjects), properties, characters, attributes, or traits (propositional predicates), and character values (Table II). In a random matrix (Table V) the element or organism is followed by a *compound proposition* consisting of many *single propositions* for which values are assigned regarding the presence and absence of characters. A *non-random* matrix is a random matrix which has been rearranged horizontally and vertically by the methods to be described below (Table VIII).

Truth tables. A truth table lists all possible combinations of values for the propositions in the table. If two-valued logic (Table IV) is used and there is one proposition, the table has two possible answers (1, 0) or $2^1 = 2$; if two propositions, four possible answers (11, 10, 01, 00) or $2^2 = 4$; if three propositions eight possible answers (111, 110, 101, 100, 011, 010, 001, 000) or $2^3 = 8$ (Table IV); and n-propositions, $2^n = 2 \times 2 \times 2 \ldots \times 2$. If three-valued logic is used and there is one proposition, the table has three possible answers (1, 2, 0) or $3^1 = 3$; if two propositions, nine possible answers (11, 12, 10, 21, 22, 20, 01, 02, 00) or $3^2 = 9$; if three propositions, 27 possible answers or $3^3 = 27$; and n-propositions, $3^n = 3 \times 3 \times 3 \ldots \times 3$; see Table IV for an example of four-valued logic.

Thus, truth tables for many-valued logic are possible but as the numbers of values and propositions increase software and hardware limitations of computers may become evident.

Horizontal non-randomization of characters in the matrix. For each character, over all organisms, count the number of times the character is present (n_1). Then, rearrange the characters in descending order of the n_1 values. The random matrix is shown in Table V; the horizontally rearranged matrix is shown in Table IV.

Vertical non-randomization of elements (organisms) in the matrix. The horizontally rearranged sequences of characters and their values for each organism are traced on a truth table with the same number of characters as the horizontally rearranged matrix (Table VII). The horizontally rearranged sequence of character values for the organism will match with one and only one pattern in the truth table. Each binary sequence of patterns in the truth table has been assigned a unique number (Table VII). The vertically rearranged matrix is shown in Table VIII.

Non-random logically summarized matrix. This represents the classification matrix and is shown in Table IX. The values for characters in each apparent cluster in this table were summarized by the logic in Table III. Briefly summarized, if a character is present (or absent) for all organisms in the cluster, the character is designated present (or absent) for the cluster or group summary. If a character is present for some organisms in the cluster and absent for other organisms in the same cluster, the character is

TRUTH TABLE ($2^8 = 256$)

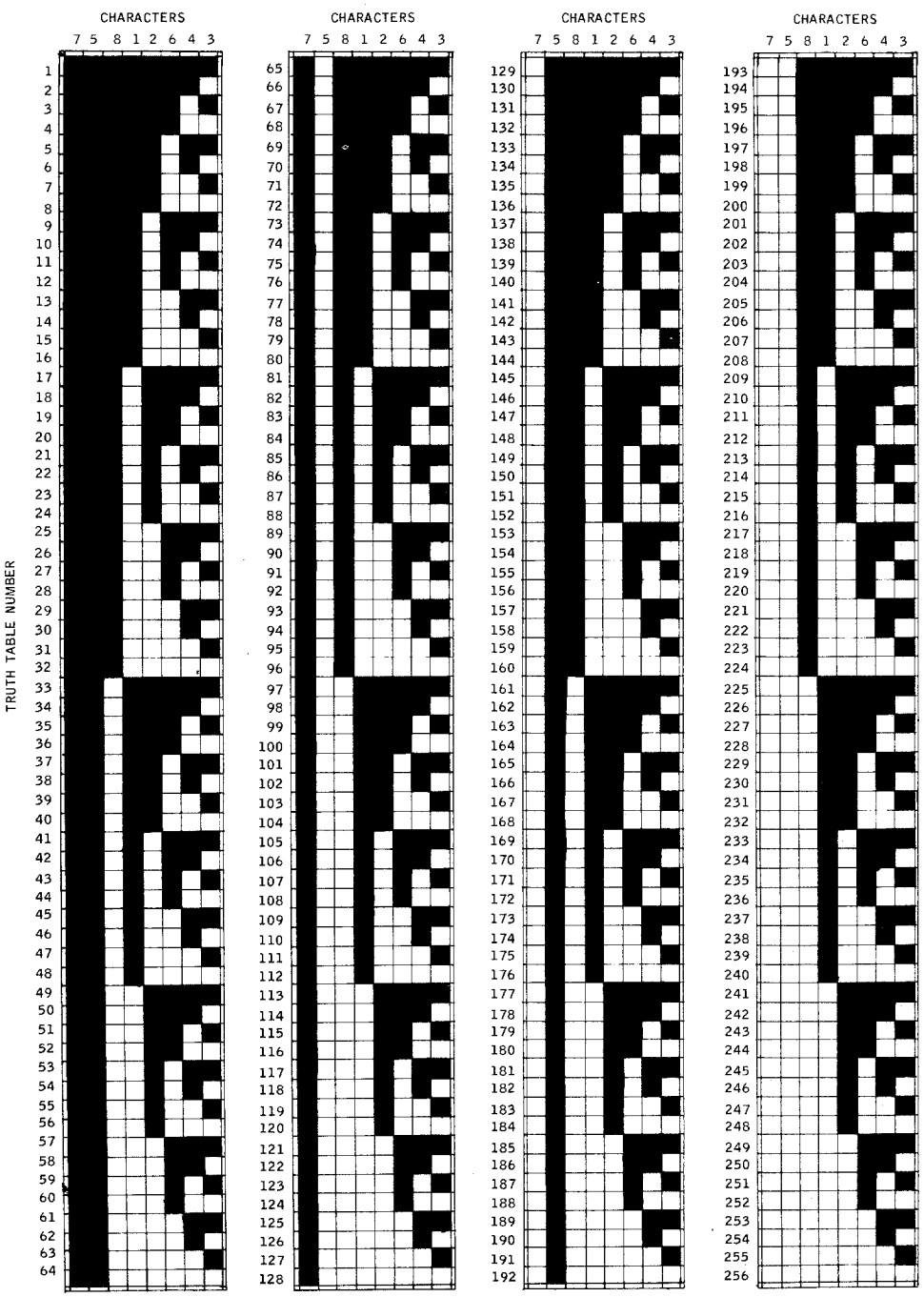
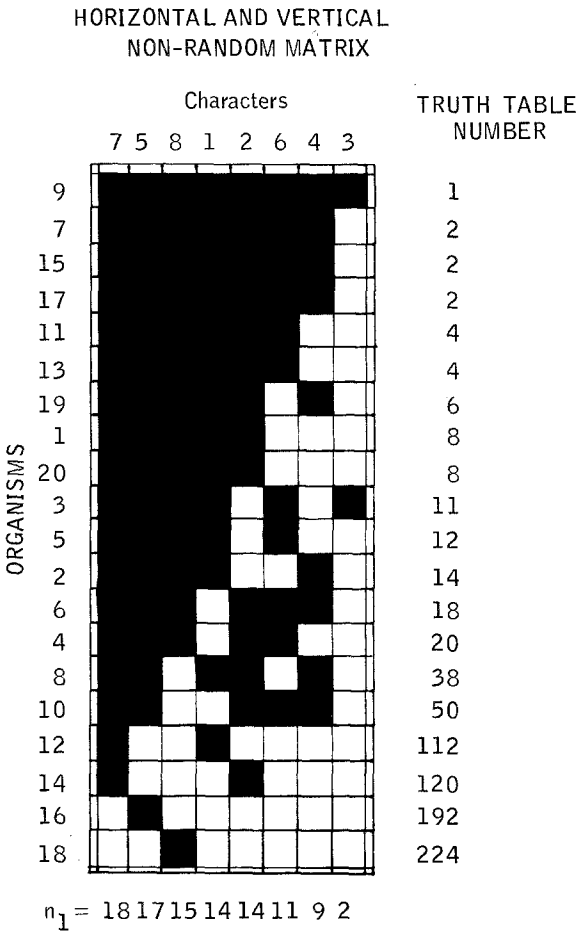


TABLE VIII

The horizontal and vertical non-random matrix obtained by first rearranging characters horizontally in descending order of their n_1 values and vertically in ascending order of the organisms' truth table numbers. Note the clusters of apparent similarity and linkage of phenotypic characters among and within organisms in the clusters or groups. Character values of each cluster are summarized by the logic in Table III. The summarized clusters or groups are shown in order in Table IX (Rypka, 1971a).



←

TABLE VII

Truth table for eight two-valued characters. Every combination of values is different and has been assigned a unique number. When a unique combination number has been found for each organism's character values, the organisms are rearranged in ascending order of the truth table combination numbers. See Table VIII which indicates combination numbers for organisms in Table VI (Rypka, 1971a).

designated variable. The classification matrix is used to determine the separation value ($S = n_1 n_0$) of Gyllenberg (1963, 1964) for each character. These values are calculated at the bottom of the classification matrix (Table IX).

S-value matrix. This is the *identification matrix* (Table X) which was obtained from the *classification matrix* by rearranging the characters in descending order of their separation values (S, calculated at bottom of Table IX) for the characters. This matrix is used to select characters for construction of an identification scheme for the organisms in the matrix. The method for construction has been explained previously (Rypka *et al.*, 1967) and will be considered only in enough detail to link identification to classification as an inverse mental activity.

An *organism versus organism* matrix is constructed and characters are selected so each character separates the most possible pairs of organisms, independent of the previously selected character(s) (Figure 2). The limitation of the method of selecting characters optimally is the limitation of not being able to make all possible combinations of n characters taken r at a time (C_r^n). The method provides for a close approximation of the shortest route to identification of organisms character-wise but not necessarily time-wise. For this reason, depending upon particular needs and situations, tests for characters that take a long time to perform, for example, gelatin liquefaction by conventional testing techniques, tests for characters with indistinct end-points (ambiguous interpretation), tests for characters requiring reagents and equipment beyond the capability of a laboratory, etc., may be excluded from being selected in the scheme. However, when the exclusion procedure for characters is used – the scheme may lose discriminatory power in terms of the number of pairs of organisms separated (Figure 2). Only *inclusive* or *exhaustive* identification schemes approach validity (Rypka *et al.*, 1967; Rypka and Babb, 1970a). For example, the finding and correlation of 'unusual microorganisms' in the disease process may not be so much changes in the particular hosts or microorganisms or both but previous incorrect identification of organisms by using exclusive or non-exhaustive schemes.

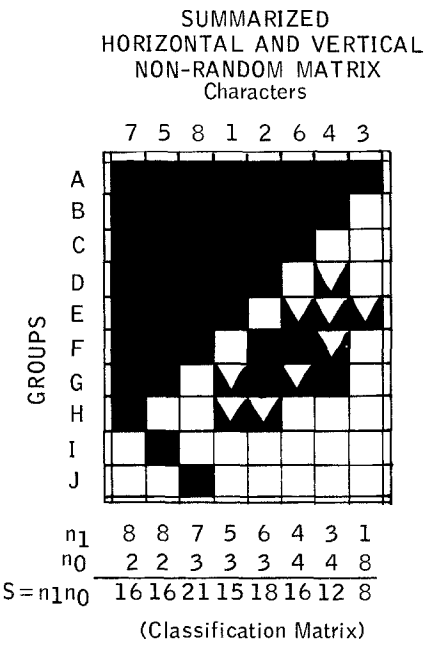
Theoretically, but not empirically, only eleven two-valued characters ($2^{11} = 2048$) would be required to identify all bacteria (Rypka, 1968) described in *Bergeys' Manual of Determinative Bacteriology* (1957). One reason for the great discrepancy between theory and experience is that there are tremendous gaps in character data for the bacteria because of the unsystematic manner in which data are collected, stored, and analyzed. If, for each organism described in the *Manual*, data were available for every different character, that is, there would be a complete organism versus character matrix with frequencies for variable traits, an optimal minimum character set could be found for purposes of identification schemes. Also, regardless of genetic, physiological, and evolutionary considerations a practical, operational classification can be made so that every organism is uniquely identifiable within an established degree of belief. One of the many contributions of numerical taxonomy is that phenotypic character data and validity of classification are increased because of increased information content (Sokal and Sneath, 1963). It is emphasized also that *Bergeys' Manual*

TABLE IX

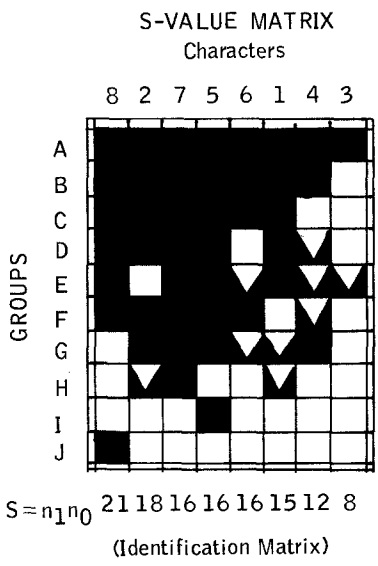
A logical summary (Table III) of Table VIII. Note each group is separable or disjoint from every other group. For each character, over all organisms, the separation value (Gyllenberg, 1963, 1964) of each cluster is calculated. This value represents the number of pairs of organisms the character will separate in an organism versus organism matrix. See Figure 2. The groups below include the logical summaries (Table III) of the following organisms from Table VIII: Group A (9), B (7, 15, 17), C (11, 13), D (19, 1, 20), E (3, 5, 2), F (6, 4), G (8, 10), H (12, 14), I (16), and J (18) (Rypka, 1971a).

TABLE X

The same as Table IX except the matrix has been rearranged in descending order of the *S*-values. This matrix may be used directly for an identification scheme or else characters may be selected (Rypka *et al.*, 1967) that optimally separate pairs of organisms. Two formats of schemes are shown in Tables XI and XII. Note the format to use really depends upon the frequency of isolation of organisms from known sources. Thus, Table XI may be the most efficient scheme if frequencies are unknown and Table XII the most efficient if frequencies are known. Identification may be approached by either strategy (See Dybowski and Franklin, 1968; Lapage, 1970). See Figure 2.



(Table IX)



(Table X)

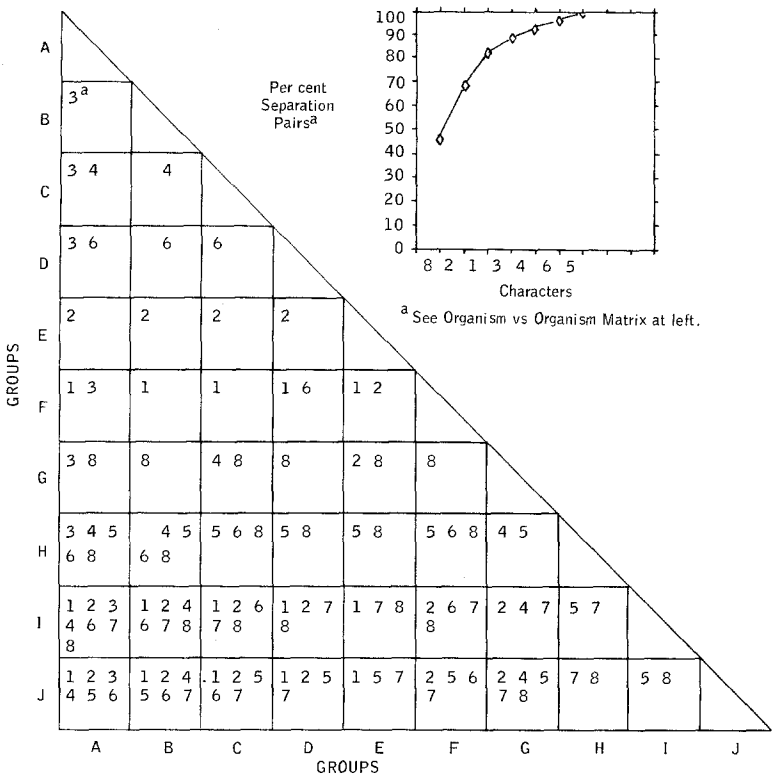


Fig. 2. An organism versus organism matrix showing which character(s) have different values for each possible pair (meaning one character = 0 for one organism and the other character = 1 for the second member of the pair). The small graph indicates the approximate per cent separation of pairs. Thus character 8 separates 21/45 (47%), character 8 and 2 together separate 32/45 (71%) characters 8, 2, and 1 together separates 37/45 (82%), etc. See Table X.

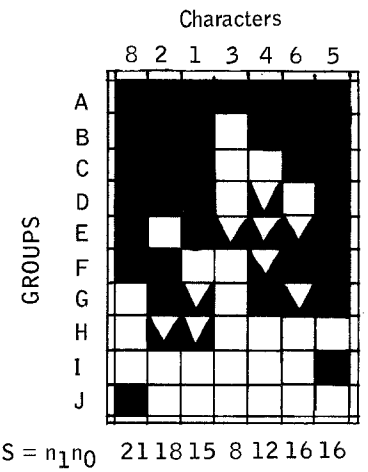
TABLE XI →

Identification scheme, matrix format. For an isolate, trace the character values for characters 8, 2, 1, 3, 4, 6, and 5. One and only one answer should be found if the isolate is on the scheme providing test sensitivity is correct.

TABLE XII

Identification scheme, truth table format. For an isolate, trace its character values for characters 8, 2, 1, 3 to find the subgroup number. For multiply answers use subgroup identification schemes. If no answer is found, and test sensitivity for characters is correct, the isolate is not on the scheme. For details see Rypka and Babb (1970a).

REDUCED S-VALUE MATRIX IN OPTIMAL CHARACTER ORDER[a]

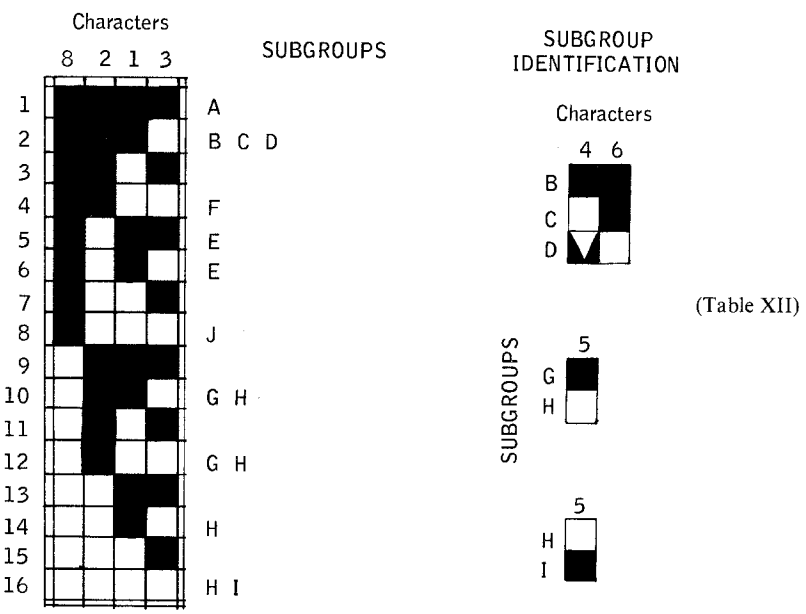


(Table XI)

[a] Means optimal order for separation of organism pairs. See Figure 2.

IDENTIFICATION SCHEME

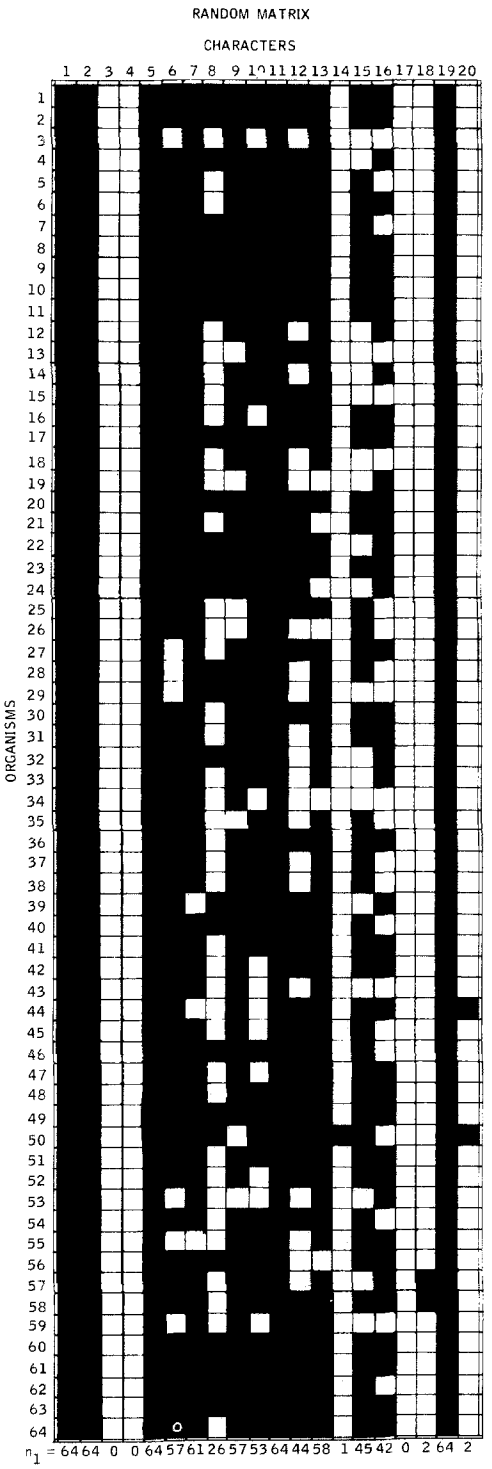
Truth Table $2^4 = 16$



(Table XII)

TABLE XIII

Random matrix of character data for staphylococci from human sources. The characters in order are: 1 (oxygen requirement), 2 (Gram stain), 3 (cell morphology), 4 (sorbitol), 5 (bile blood agar, 10%, grow), 6 (oxidative – fermentative equal fermentative), 7 (haemolysis, beta), 8 (mannitol), 9 (methylene blue, 0.1%, reduce), 10 (arginine, produce ammonia from), 11 (catalase), 12 (acetylmethylcarbinol), 13 (phosphatase), 14 (arabinose), 15 (lastose), 16 (glycerol), 17 (transparent, colony), 18 (red, colony), 19 (bile blood agar, 40%, grow), 20 (cellobiose).



of Determinative Bacteriology (1957) is an excellent and monumental attempt to bring bacterial character data into systematic order.

The methods of testing for bacterial characters explained in the next section may be found in Cowan and Steel (1965) and Harrigan and McCance (1966) except for the base medium for detecting acid production from carbohydrates. This medium, designated TPP3 for tryptose and proteose-peptone No. 3, was a modification of that of McDade and Weaver (1959). The composition of TPP3 was tryptose 10 g, proteose-peptone No. 3 (Difco) 10 g, sodium chloride, 5 g; KH_2PO_4, 0.6 g; agar, 0.6 g; phenol red, 0.054 g; distilled water 900 ml, adjust to pH 7.3 – 7.4. To each 2.7 ml of medium in 12 × 125 mm tube was added 0.3 ml of the Seitz sterilized 10% carbohydrate. All cultures were checked for purity by the Gram staining procedure. The characters studied are given in the caption of Table XIII.

3. Results

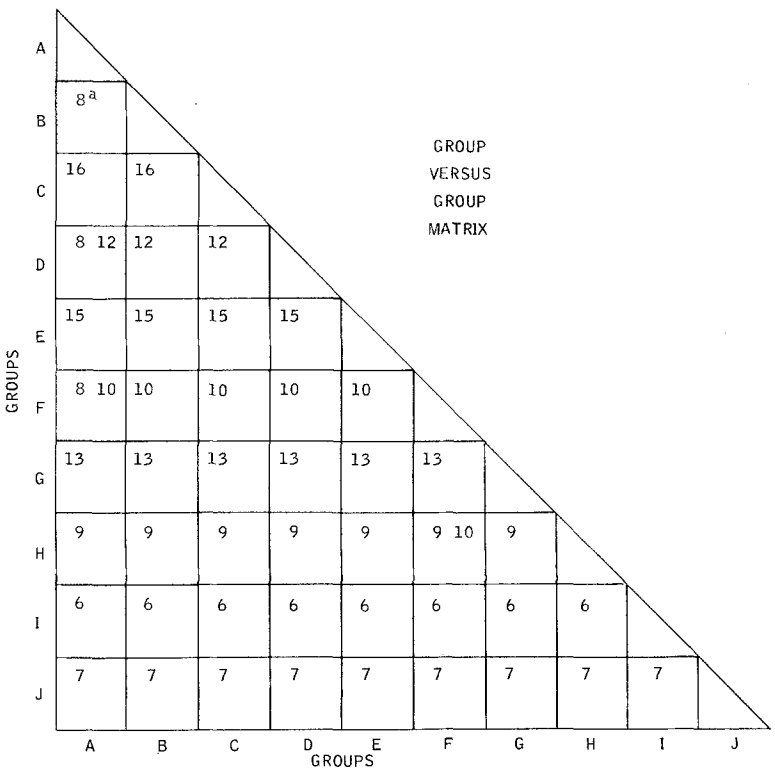
To illustrate continuous or truth table classification and identification, data were accumulated continuously for twenty characters of heterotrophic, aerobic, catalase positive, Gram positive cocci isolated from human sources. The character data were recorded by the standard two-valued method of character present = 1 and absent = 0. These data were processed by batch processing using programs written by Robert Babb in Burroughs ALGOL 60 or by time-sharing using a teletype. The latter programs were written in COBAL at the Computer Section.

The *random matrix* of 64 isolates and 20 characters is shown in Table XIII. Horizontal non-randomization is shown in Table XIV. The vertical non-randomization of the matrix is shown in Table XV and ten clusters of apparent similarity are discernible. The truth table number for each organism is given at the right of the table.

For twenty, two-valued characters, the size of the truth table to determine the truth table number would be $2^{20} = 1\,048\,576$. The actual truth table number of the isolate is determined by calculation and it is not necessary to construct a truth table of this size. The importance of this is in initial taxonomic studies, e.g., a numerical taxonomic study, where a given set of characters are studied for many isolates, each isolate will have a pattern of character values which may be assigned a unique number and, if necessity requires it, actual disjoint clusters may be obtained continually or continuously as the study progresses. Truth tables of the size 2^{99} have been used by this method and vertical rearrangement of matrices accomplished to obtain apparent clusters.

The truth table numbers for the 64 isolates vary from 64 to 20 864. The reason for the narrow range of numbers, 64 to 20 864 in 1 048 576, is that a group was delimited to begin with in the twenty characters, i.e., heterotrophic, aerobic, catalase positive, Gram positive cocci from human sources.

The classification matrix is shown in Table XVI. The group versus group matrix is shown in Figure 3 and was constructed by the same method discussed for Figure 2. An identification matrix is shown in Table XVII with characters not arranged in



[a] Means character 8 is different for each member of the pair. Therefore, each organism is disjoint or separable from the other.

Fig. 3. An organism versus organism matrix showing which character(s) have different values for each possible pair (meaning one character = 0 for one organism and the other character = 1 for the second member of the pair). See Table XVI.

TABLE XIV

Horizontal matrix of Table XIII. See Table XIII for character codes.

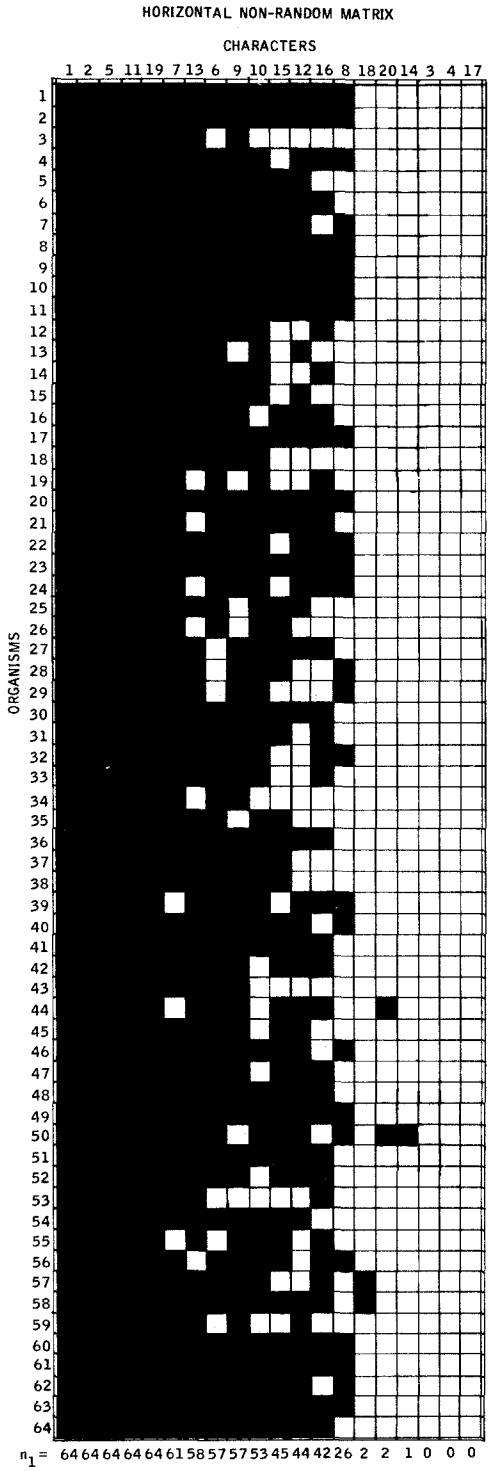


TABLE XV

Horizontal and vertical matrix for Table XIII. See Table XIII for character codes.

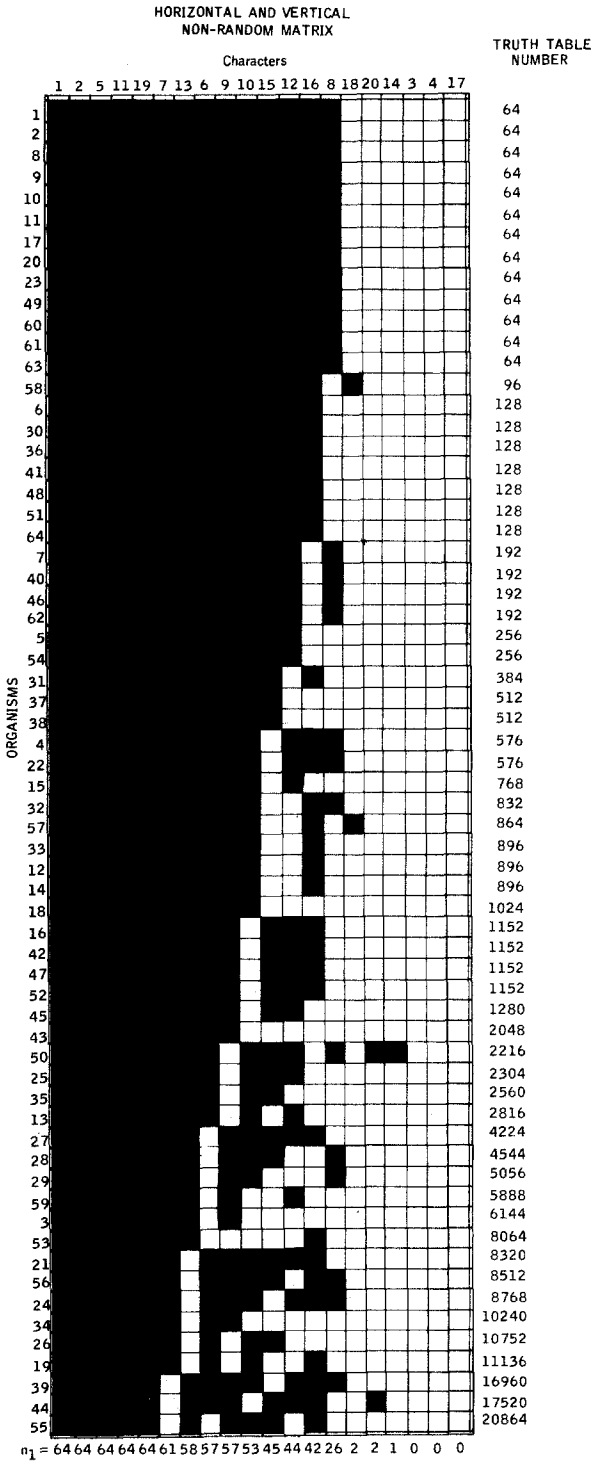
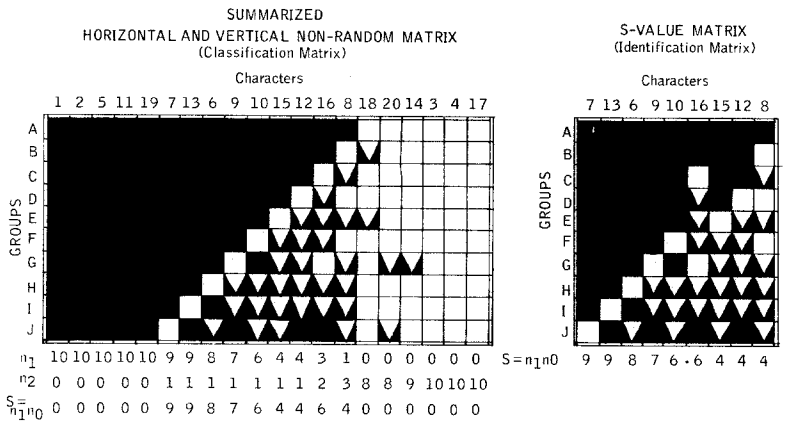


TABLE XVI and TABLE XVII

Table XVI. Classification matrix for Table XV. The groups below include the logical summaries (Table III) of the following organisms from Table XV: Group A (1, 2, 8, 9, 10, 11, 17, 20, 23, 49, 60, 61, 63), B (58, 6, 30, 36, 41, 48, 51, 64), C (7, 40, 46, 62, 5, 54), D (31, 37, 38), E (4, 22, 15, 32, 57, 33, 12, 14, 18), F (16, 42, 47, 52, 45, 43), G (50, 25, 35, 13), H (27, 28, 29, 59, 3, 53), I (21, 56, 24, 34, 26, 19), J (39, 44, 55). See Figure 3. Table XVII. The *S*-value matrix for Table XVI. For complete separation of pairs every character is required. From this matrix identification schemes in the matrix format (Table XI) or truth table format (Table XII) could be constructed.



descending order of their rate of separation of pairs because in Figure 2, all characters in the matrix are required for separation. Selection of characters could be done (Table IX) and schemes similar to those in Tables XI and XII constructed. Performing tests for characters in this table in the order listed provides a nearly optimal route to identification of an isolate. Mention should be made that identification schemes for each subgroup in Table I have been constructed and also smaller subschemes of these which are used on the teletype and time-sharing. In addition, if an investigator does not want to follow the shortest route or has only limited values for character data, each large matrix (Table I) is easily searched automatically and all possible organism answers obtained on the basis of character value information available.

4. Discussion

Classifications are dynamic, not static; tentative, not final, simply because all elements to be classified usually are unavailable for study at one time. A method has been developed that would continuously (meaning more than continually and conceivably less than continuous) classify elements at a high level of similarity and yet create disjoint groups. The method must – and does – readjust classification of elements with time recognizing the dynamic state of nature. The method is analogous to finding an instantaneous rate of change of a function. In the present work the objective is to find an instantaneous correlation of character position (pattern) of elements (organisms) with time. This was done using the concept of the 'logical alphabet' of Jevons (1877)

which linked classification, based upon similarities, and identification based upon differences.

Confusions exists regarding the inferential basis for classification. White (1937) considered classification as both deductive inference ('analytic') and inductive inference ('constructive'). Stanier (1970) believes that 'traditional phylogenetic taxonomies' are deductive and that genetically based classification will have an inductive basis. Steel (1962) considered taxonomy an art, identification a science. Jevons (1877) considered classification was a science and identification deductive logic. Problems relating to classification and identification become a matter of analytic philosophy (Pap, 1958) and the way out of existing problems may be the method of explication of Carnap (1950) in which concepts in vague language are restated in more precise language. Thus, stated explicitly: classification is an inductive mental activity and identification is a deductive mental activity (Rypka, 1970d). The two are inverse mental activities (Jevons, 1877; Mayr, 1969) and deductive-inductive methods in science are well-known (Mills, 1850; Venn, 1889). On this basis it is recognized readily that biological classification is the result of imperfect induction, that is, imperfect because of lack of sufficient data to sort accurately the elements into groups on the basis of similar characteristics. Generally, more than the facts are asserted in any classification and therefore passage is made from perfect to imperfect induction.

The method described for continuous or truth table classification is merely a device. It is meant for rapidly collecting and analyzing patterns of data about microbes. The method does have distinct advantages; (1) Rapidity. By either batch processing or time-sharing the method is very rapid. (2) The method may be used within existing systematic and nomenclatural rules and results may be readily compared and correlated with existing classifications. (3) The method would be excellent for determining epidemological patterns and for further statistical analysis to indicate shifting flora patterns, for example, in closed environments such as spacecraft. (4) Nakabayashi (1971) has suggested analogous methods for monitoring patients in intensive care units. (5) The method is generalized and applicable to other logical classification and identification problems.

Sneath (1968), in discussing the future outlines of bacterial classification, mentions the 'fusion' of strategies for classification and for identification. The simple, practical approach discussed in the present work fuses both classification and identification to a truth table and considers them as inverse mental activities. For approaches to these problems interesting advances may be found in the works of Gyllenberg (1965; 1967a, b), Niemela *et al.* (1968), Dybowski and Franklin (1968), Morse *et al.* (1968), Lapage *et al.* (1970), and Pankhurst and Walters (1970). For suggested methods of coding microbial data see Rypka and Babb (1970a). For excellent and extensive coding information see Rogosa *et al.* (1971).

5. Conclusion

The method of continuous or truth table classification and identification is an ade-

quate, practical, rapid, interim method of data collection and analysis. Classification and identification are considered inverse mental activities of imperfect induction and deduction, respectively. The method is generalized and applicable to other classification and identification problems.

Acknowledgements

To: Richard Goodrich, Computer Systems Manager, Lovelace Foundation, for developing the method for time-sharing and for COBAL programs; Robert Babb for ALGOL programs; Rosemary Speeker Rypka and Pauli Paulikonis, assistance in the preparation of the manuscript. Donnell Hester and O. B. Weeks, William Beaumont Army Hospital, El Paso, Texas and Research Center, New Mexico State University were kind enough to engage in discussions concerning this method and to loan their numerical taxonomic data for analysis by this method. They will publish their findings later.

References

Breed, R. S., Murray, E. G. D., and Smith, N. R.: 1957, *Bergey's Manual of Determinative Bacteriology*, 7th edition, Williams and Wilkins Company, Baltimore.

Carnap, R.: 1950, *Logical Foundations of Probability*, University of Chicago, Chicago.

Caws, P.: 1965, *The Philosophy of Science, A Systematic Account*, D. van Nostrand Company, Inc., Princeton.

Cowan, S. T. and Steel, K. J.: 1965, *Manual for the Identification of Medical Bacteria*, Cambridge University Press, London.

Dybowski, W. and Franklin, D. A.: 1968, *J. Gen. Microbiol.* **54**, 215–29.

Gyllenberg, H. G.: 1963, *Ann. Acad. Sci. Fennicae Ser. A. IV. Biologica* **69**, 1–23.

Gyllenberg, H. G.: 1964, *Ann. Acad. Sci. Fennicae Ser. A. IV. Biologica* **81**, 1–23.

Gyllenberg, H. G.: 1965, *J. Gen. Microbiol.* **39**, 401–5.

Gyllenberg, H. G. and Eklund, E.: 1967a, *Ann. Acad. Sci. Fennicae Ser. A. IV. Biologica* **113**, 3–19.

Gyllenberg, H. G., Woznicka, W., and Kurylowicz, W.: 1967b, *Ann. Acad. Sci. Fennicae Ser. A. IV. Biologica* **114**, 3–15.

Harrigan, W. F. and McCance, M. E.: 1966, *Laboratory Methods in Microbiology*, Academic Press, New York.

Hayakawa, S. I.: 1964, *Language in Thought and Action*, Harcourt, Brace, and World, Inc., New York.

Hilbert, D. and Ackermann, W.: 1950, *Principles of Mathematical Logic*, Chelsea Publishing Company, New York.

Hunt, E. B.: 1962, *Concept Learning. An Information Processing Problem*, John Wiley and Sons, Inc., New York.

Jevons, W. S.: 1877, *The Principles of Science. A Treatise on Logic and the Scientific Method*, Dover Publications, Inc., New York.

Lapage, S. P., Shoshana, B., Willcox, W. R., and Curtis, M. A.: 1970, 'Computer Identification of Bacteria', in *Automation, Mechanization and Data Handling in Microbiology* (ed. by A. Baillie and R. J. Gilbert), Academic Press, London, pp. 1–22.

Lukasiewicz, J.: 1920, *Ruch Filozoficzny* (Lwow) **5**, 169–71.

Mayr, E.: 1969, *Principles of Systematic Zoology*, McGraw-Hill Book Company, New York.

McDade, J. J. and Weaver, R. H.: 1959, *J. Bacteriol.* **77**, 65–9.

Mills, J. S.: 1850, *A System of Logic, Ratiocinative and Inductive*, Harper and Brothers, Publishers, New York.

Morse, L. E., Beaman, J. H., and Shetler, S. G.: 1968, *Taxon* **17**, 479–83.

Nakabayashi, K.: 1971, personal communication. See also: Nakabayashi, K.: 1970, *J. Japan. Anesthesiol. Soc.* **19**, 854–68.

Niemela, S. I., Hopkins, J. W., and Quadling, C.: 1968, *Can. J. Microbiol.* **14**, 271–78.
Pankhurst, R. J. and Walters, S. M.: 1970, '14. Generation of Keys by Computer', in *Data Processing in Biology and Geology* (ed. by J. L. Cutbill), pp. 189–203, Systematics Association Special Volume No. 3.
Pap, A.: 1958, *Semantics and Necessary Truth. An Inquiry into the Foundations of Analytic Philosophy*, Yale University Press, New Haven.
Rogosa, M., Krichevsky, M. I., and Colwell, R. R.: 1971, *Intern. J. System. Bacteriol.* **21**, 1A–175A.
Rosser, J. B. and Turquette, A. R.: 1952, *Many-valued Logics*, North-Holland Publishing Company, Amsterdam.
Rypka, E. W.: 1968, 'Use of Combinatorial Analysis for Constructing Bacterial Identification Schemes', in *Research and Development in Progress – Biology and Medicine*, United States Atomic Energy Commission, Division of Technical Information Extension, Oak Ridge, Tennessee, p. 61.
Rypka, E. W.: 1970, 'Classification and Identification as Problems in Induction and Deduction', Joint Meeting Arizona – Colorado – New Mexico Branches, American Society for Microbiology, University of Arizona, Tuscon, Arizona, November 13–14, 1970.
Rypka, E. W.: 1971a, 'Continuous Classification and Identification', *Abstracts*, 47th Annual Meeting, Southwestern and Rocky Mountain Division, American Association for the Advancement of Science, Arizona State University, Arizona. April 21–24, 1971, p. 82.
Rypka, E. W.: 1971b, 'Automatic Identification, Data Accumulation, Analysis, and Reclassification', *Bacteriological Proceedings*, 71st Annual Meeting, American Society for Microbiology, Minneapolis, Minnesota, May 2–7, 1971, 37.
Rypka, E. W. and Babb, R.: 1970a, *Med. Res. Eng.* **9**, 9–19.
Rypka, E. W. and Babb, R.: 1970b, 'Automated Data Reduction Using Octamerous Identification Schemes', *Bacteriological Proceedings*, 70th Annual Meeting, American Society for Microbiology, Boston, Mass. April 26–May 1, 1970, 93.
Rypka, E. W. and Babb, R.: 1970c, 'Automatic Construction and Use of Inclusive Identification Schemes', *Abstracts*, Xth International Congress of Microbiology, Mexico, D.F., August 9–14, 1970, 191.
Rypka, E. W., Clapper, W. E., Bowen, I. G., and Babb, R.: 1967, *J. Gen. Microbiol.* **46**, 407–24.
Sneath, P. H. A.: 1968, *Class. Soc. Bull.* **1**, 28–45.
Sokal, R. R. and Sneath, P. H. A.: 1963, *Principles of Numerical Taxonomy*, W. H. Freeman and Company, San Francisco.
Stanier, R. Y.: 1970, 'Towards an Evolutionary Taxonomy of Bacteria', for *Recent Progress in Microbiology*. **X**, Xth International Congress for Microbiology, Mexico, D. F., August 9–14, 1970.
Steel, K. J.: 1962, 'The Practice of Bacterial Identification', in *Microbial Classification* (d. by G. C. Ainsworth and P. H. A. Sneath), Cambridge University Press, Cambridge, pp. 405–32.
Styazhkin, N. I.: 1969, *History of Mathematical Logic from Leibniz to Peano*, The M.I.T. Press, Cambridge.
Venn, J.: 1889, *The Principles of Empirical or Inductive Logic*, MacMillan and Company, New York.
White, P. B.: 1937, *Zbl. Bakt.* II Abt. **96**, 145–9.
Zermelo, E.: 1908, 'Investigations in the Foundations of Set Theory I', in *From Frege to Godel, A Source Book in Mathematical Logic*, 1879–1931, by van Heijenoort, J.: 1967, Harvard University Press, Cambridge, pp. 199–215.