Results of the Application of Taximetric Techniques in the Study of Varieties of Triticum

by

Allan McKeag Hale

Outline

I. Introduction

A. History and Background

B. Purpose of Experiment

1. Interspecific Relationship

2. Arrangement of Groups

II. Character Analysis

A. Character Selection

B. Morphological Significance

C. Weighting of Characters

D. Special State Assignment

III. Clustering

A. Character Arrangement for Classification

B. Clustering Map

1. Groups

2. Internal Connections

3. Articulators

4. Host Values

IV. Skyline

A. Introduction of Objects

B. G-value Importance

C. Group significance

D. Level and its Importance

E. Rye as a Check

V. Considerations

A. Chromosomes

1. Ploidy Level

2. Grouping

B. Evolutionary Significance

C. Rye Significance

D. Conclusions

The particular experiment undertaken in this study was the determination of the degree of interrelatedness of the stocks within the Triticum species of cereal grain. An attempt was made through the use of computer techniques to sort out and emphasize the valuable information which was relevant to the study by way of describing the particular characteristics sought with accuracy. I will in this paper strive to discuss several areas of taximetrics as they relate to my individual study.

The purpose in my study was not to create an arbitrary taxon, for within a single species this is an uncalled for project. A preliminary investigation was made to find what natural progression or simularity was to be found or displayed in the Triticum samples which were selected. Before a consideration of the results of the study will be made, a little background material concerning Triticum should be contemplated. Wheat is an extremely inbred stock, that is, the individual plant is usually self fertilizing. Wheat accomplishes its self fertilization while the sexual organs or parts are still covered with the sheath or "boot" of the plant. Due to the nature of this variety of fertilization the wheat plant is found to exhibit the genetic characteristics of a highly inbred strain. It should be noted though, that wheat is not completely inbred and is found to have a naturally occuring outbreeding pattern of from 1 to3 percent in a homogeneous and arbitrarily selected population. This outbreeding is the reason for the hybrid stocks of wheat which are now under intensive cultivation by man. The outbreeding programs are primarily developed for two reasons. First and foremost these programs are undertaken to produce wheats with greater grain yield either by larger or by more numerous seeds. The second reason for the hybrid programs is to develop wheats with more adaptable characters for a particular climatic condition. This adaptability can be understood as either genetic selection for a particular climate or the development of a resistance within the wheat to climatic conditions or disease which occurs within the particular variety's ecological range. For purposes of this experiment I was fortunate enough to obtain some sixty-eight different varieties of wheat for the program. Roughly, the samples can be divided into three main categories or groups. First, there is the Regional Hard Red Spring wheat variety which included twenty-four of the samples in the collection. Second, the State Spring wheat variety included twenty different strains of this stock. Third, the last group for the collection was the dwarf Winter wheat stock which included twelve different types for the study. The designation of a spring wheat group means literally that the group was planted in the spring and came up during the summer of that same year. The winter wheat stock is descriptive of a group planted late in the fall of a year in which the wheat was allowed to vernalize (age during a freezing period) for that winter and the stock came up in the spring of the following year. This difference develops certain morphological differences in the wheat but is primarily an economic factor for the maximum use of the available land, for crop propagation, and has only segregatory significance to this study.

For the characters which were needed for the program I selected some gross morphological characteristics which were common to all of the samples. Such factors as the height of the plant, flower arrangement, number of flowers, spikelet length, spikelet diameter, brush length, angle of reflection of the brushes, length of the seeds, colour of the seeds, as well as the degree of ease by which the seeds could be removed from the bracts of the plants, formed the bases of the charanal program (reference). Several of the characters I discovered later had increased weight in the program. This weighting was unintentional and in a later run the stress was taken off of the importance of some of these. An example of this unintentional weighting was, where the brush length and the reflection of the brush were considered as two distinct characters. In the characters there were several states including the "not present" state which in the case of the brush length told all of the information which was necessary for the reflection state and here, there was a redundancy of information. For, if there were no brush length, the brush was not present and then for the reflection character the state would automatically not be present. It turned out that this particular redundancy was not deleterious and the overlap in information was acceptable for the study.

While charanal offered a means of determining those characters which offered the most usable types of information to the wheat study, another facet of the program, clustering, combined the advantages of the charanal system with the added feature of grouping the objects within the study according to their overall similarity considering all of the characters and the states of those characters which the objects exhibited. By the use of the clustering program a map of the objects was made. The clustering information provided data concerning the objects which at different levels of similarity were most alike. In this way a map of the clusters formed by the computer using the charanal information provided, was formed. These clusters were of prime importance in the determination of which objects or samples were categorically most similar. The map not only showed the cluster membership but also how similar objects were related within the individual cluster and which ones were the more important members of each cluster. An interesting feature of this clustering program is that the members of the study which join the major clusters, the articulators in the case of my study were very conspicuous by their absence. This I found later was the normal state of affairs with the studies which include only one species. A small number of articulators characterize the intraspecific clustering type, whereas the larger number of cluster articulators are indicators of an interspecific clustering program. The moat values within my study fell within the limits of 0.1 to 0.01 . The moat is the mathematical distance which separates one level of similarity from another.

In the clustering program there is a portion which is devoted to the graphic relation of the objects in the groups of a particular study and their value of similarity. This graphical representation is the skyline plot. The skyline plotter organized the objects of my study into those ob jects which were most similar and placed them on the horizontal axis of the graph and plo ted against them was the C-value, the value of similarity. The C-value is usually a graduation from 1.00 to the lowest C-value which is relevant to the study under consideration. In my particular case the C-value ran from 1.00 to0.40 . In this study I had two objects which were identical in all characteristics yet they were of different varieties and they were found to have a C-value of 1.00 . I found that as the C-value decreased the addition of more objects took place at a rapid rate. At each addition of a new member or members to the groups present on the skyline, the addition was classed a level in the study. In my particular instance at level nine there were six groups of objects present and the overall significance of this was that each of the clusters was a particular strain within the species. For example, there was a cluster including all of the Dwarf Winter wheat varieties in the study. Yet another cluster included the majority of the Regional Hard Red variety. It was interesting to note that all of the winter wheats were clustered at a higher C-value in the skyline than the spring varieties indicating that the winter wheats were more closely related than the other strains. I will consider the evolutionary significance of this fact later in this paper. In the study were included several rye samples which later in the study acted as checks on the clustering method. These rye samples were always found to be grouped together in the clustering program. In only one case did a rye sample fall into another classification than its own seperate group. It was found to be of use to have the rye in the study for it offered a guidestick as to the behaviour of the other objects in the study.

For example, was there missing information which covered this?

A factor which bothered me in the wheat study was the fact that at level nine there were six groups present and they remained as six distinct groups for quite a while considering the C-value as a basis of similarity. I mulled this fact over for almost a week till I hit on the idea that perhaps there was some basic genetic reason for this clu ster seperation. Looking at a chromosom e atlas I immediately discovered that the six seperate groups of wheat stock were the representation of the different ploidy levels of the wheat present in the study. I had not considered the ploidy level in my characters for comparison but the fact remains, that the ploidy levels were printed out vividly in the cluster map and the skyline plot. At level nine the six ploidy levels and the objects in the groups were:

Cluster N o.	Ploidy Level	Objects Included
1	14	101 148
2	21	104 121 118 117 204 143 135 205 203 108 212 134 202 220 136 201 208 136 207 206 211 217 209 113 216 122 141
3	28	119 132 147 142
4	42	106 145
5	56	105 109
6	63 (84)	114 131 138
7	(RYE OBJECTS)	

Apparently these objects within the program had characters which due to the morphology were in some manner related to their chromosome number. No doubt the factor involves several of the physical aspects of the wheat plants.

The clustering program supports also the evolutionary significance of the cluster arrangement. The most early members of the wheat stock which were plotted on the skyline were the winter stock. This is one of the oldest strains of wheat which has been cultivated by man and due to the intensive inbreeding of the wheat through time, the similarities of the winter wheats as a group is very great. The spring wheats on the other hand are more recent in their development and are primarily composed of the outbred and the artificially bred hybrids and are newer in the evolutionary sense. This information was supported in the cluster analysis and also in the skyline plot which showed the winter wheats displayed as a cluster sooner at a higher C-value than the dwarf and the spring stocks which clustered later and in that order on the skyline. Also, from the cluster analysis I noted that there were a great deal more internal connections within the winter variety than there were in the dwarf and the spring varieties. This also supports the theory that the winter variety being a more highly inbred stock would have a higher degree of relatedness within the strain than the new hybrids which are once outbred and then remain in that state with little or no cross pollenation among different varieties of the spring stock.

The total significance of the rye samples in the study was as I mentioned earlier that they were a check on the method accuracy and provided a measure of the ability of the computer technique to sort out exceptional objects. Rye was however, combined with the wheat in the study at a relatively high C-value possibly because it is the most close relative to the wheat of all of the species of Gramineae.

Briefly, in conclusion I would like to state that I have in part accomplished what I set out to do and that was to determine if any structuring was in fact to be found in the wheat species. I would however like to undertake a more rigorous program of study with wheat (now that my curiosity has been stimulated) with additional patterns of character analysis included in it. For this work I also would like to obtain many more samples than the meagre lot which I had to deal with in this present study. I have learned much by was of the theory and application as well as the consideration of some of the philosophical questions which exist in taximetrics. I intend to use the information and technique which I have learned in this course in taximetrics to aid my own independent study. I am analyzing some of the cereal grains and the taximetric methods are very relevant to this study, and are a valuable tool for the biologist.

Very good work, Allen. I, too, would like to see a continuation of this work. I think it would be wise if we invited Lorraine to share with them the results of your work. Let's talk about it!

Taxonomic Classification of Populations of *Sceloporus undulatus erythrocheilus* Maslin (Reptilia, Iguanidae).

John W. Ferner

TABLE OF CONTENTS

INTRODUCTION

Sceloporus undulatus erythrocheilus, as described by Maslin (1956), occurs exclusively in rock habitats along the eastern foothills of the Rocky Mountains. Populations of this lizard occurring in suitable rock habitats often appear to be separated from other populations by several possible types of geographical barriers (canyons, streams, arroyos, and so on). The intent of this study was to attempt to find a taxonomic classification of lizards selected from four different populations. The purpose of this study was also to learn and employ various taximetric and numerical taxonomic techniques.

MATERIALS AND METHODS

Specimens of <u>Sceloporus</u> <u>undulatus</u> <u>erythrocheilus</u> from four locations were obtained from the personal collection of John Ferner (coded JWF) and from the University of Colorado Museum collection (coded UCM). Five adult males and five adult females were selected from each location, making a total of 40 specimens in the study. The collection numbers of each specimen are given in Table 1. Fig. 1 shows the approximate locations of the four populations investigated: 1). Boulder County, north of Boulder Creek; 2). Boulder County, Flagstaff Mtn.; 3). Fremont County, 10 miles east of Canon City; and 4). Baca County, Cottonwood Creek.

Characters for classification were chosen and the state of each character was recorded on a 3"X5" card for each specimen. Any character which had more than one definable state was acceptable for the study. The characters chosen and their influence on the study will be mentioned in the next section.

All information gathered from the specimens was analyzed with the computer programs of the Taximetrics Lab at the University of Colorado. The character analysis program is as described by Estabrook (1967) and provides a measure of interdependence of the descriptors. The characters then selected for the classification were used

Table 1.

List of the specimens used in this study, including the character states used in the Dec. 19 run.

Collection #	Study #	Collection Site	Sex	States of Characters as given in Table 3.
UCM 14898	1	Baca Co.	M	2212323421 53
UCM 14897	2	Baca Co.	M	1222323231 42
UCM 14892	3	Baca Co.	M	2223223531 43
UCM 14903	4	Baca Co.	M	2223123252 52
UCM 14904	5	Baca Co.	M	2221222312 42
UCM 5680	6	Fremont Co.	M	2221222342 13
UCM 5686	7	Fremont Co.	M	3111112121 22
UCM 5665	8	Fremont Co.	M	2111112431 53
UCM 5677	9	Fremont Co.	M	2211111251 42
UCM 5683	10	Fremont Co.	M	1122122231 22
JWF 81	11	Boulder S.	M	1221111221 22
JWF 41	12	Boulder S.	M	2111111252 42
JWF 85	13	Boulder S.	M	2113112231 31
JWF 45	14	Boulder S.	M	1121111121 46
JWF 66	15	Boulder S.	M	1111212222 43
JWF 28	16	Boulder N.	M	1114231322 43
JWF 87	17	Boulder N.	M	1111211222 32
JWF 2	18	Boulder N.	M	2121111122 42
JWF 99	19	Boulder N.	M	2221111142 43
JWF 11	20	Boulder N.	M	2111111121 23
UCM 14907	21	Baca Co.	F	2221242432 62
UCM 14891	22	Baca Co.	F	2222222211 23
UCM 14910	23	Baca Co.	F	2111123212 33
UCM 14905	24	Baca Co.	F	3221221431 42
UCM 14909	25	Baca Co.	F	2221243231 33
UCM 5667	26	Fremont Co.	F	2223213132 23
UCM 5669	27	Fremont Co.	F	2213242241 33
UCM 5666	28	Fremont Co.	F	2113222241 33
UCM 5661	29	Fremont Co.	F	2221242241 33
UCM 5682	30	Fremont Co.	F	3222112111 22
JWF 50	31	Boulder S.	F	3124111121 43
JWF 48	32	Boulder S.	F	1123143132 52
JWF 19	33	Boulder S.	F	2113111352 43
JWF 64	34	Boulder S.	F	3111211142 23
JWF 79	35	Boulder S.	F	2223242232 43
JWF 92	36	Boulder N.	F	3223111222 43
JWF 72	37	Boulder N.	F	3111111132 33
JWF 1	38	Boulder N.	F	2221212231 43
JWF 68	39	Boulder N.	F	2223112452 53
JWF 74	40	Boulder N.	F	3113111222 54

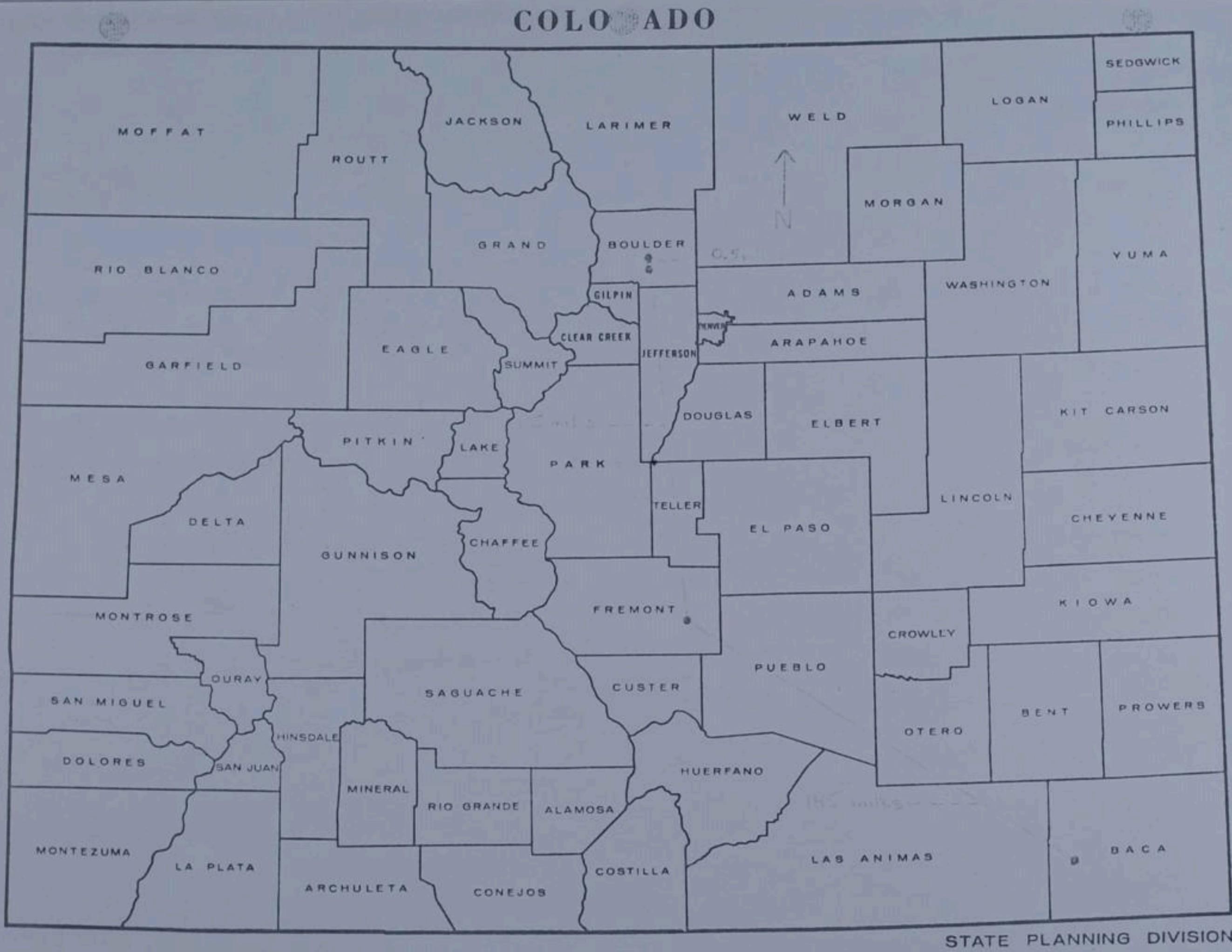


Fig. 1. Location of populations used in this study.

to compute a similarity measure for each pair of speci-
mens.(Estabrook and Rogers, 1966). Wirth, Estabrook, and Rogers (1966) and Estabrook (1966) explain the graph theroy model used to group the specimens into clusters and create a skyline plot. Subgraphs were drawn from the results on the computer print-out in order to visualize the clusters and their interrelations. For a detailed explanation of the taximetric principles used consult the above mentioned publications.

OBSERVATIONS

After selecting the specimens to be studied, the characters to be used in the classification were chosen. In the first character analysis several items were considered as possible influencing factors. Table 2 lists the first 10 characters; sex, time of year collected, and location were included in order to discover their redun-

Table 2.

Characters used in the first analysis.

Character and Number	Character State and Number
1. Sex	1. Male 2. Female
2. Snout-vent length	1. Small (55-60 mm) 2. Medium (61-67 mm) 3. Large (68-73 mm)
3. Time of year collected	1. May 2. June 3. July 4. August 5. September
4. Frontoparietal Left	1. Whole 2. Split
5. Frontoparietal Right	1. Whole 2. Split
6. Gular patches	1. Broadly joined 2. Joined, but incompletely 3. Just joined (by one scale) 4. Just separate 5. Separate and reduced
7. Ventral spotting	1. Heavy 2. Light 3. None
8. Abdominal patch shape	1. Square 2. Rounded 3. Shape indeterminable 4. Pigment lacking
9. Location	1. Boulder County, north 2. Boulder County, south 3. Fremont County 4. Baca County

Table 2. (cont.)

10. Front profile	1. Crest high 2. Crest medium 3. Crest low

dancy with other characters, not to be used as characters themselves. The only significant similarity encountered was that sex is closely related to the joining of gular patches. An attempt was made to factor sex out of the character which resulted in character 4 of Table 3. Conditional probability tables were provided by the computer and show some interesting relationships between characters, see examples in Table 4 through Table 8. Most of the significance of these seems to be the relationship between characters and location. The character states for each specimen are listed in Table 1.

The first 7 characters in Table 3 were then run through the clustering program as simple characters. In a second run through the clustering program, characters 8 through 12 in Table 3 were added to the study. A character analysis of these was not available from the computer, but a manual check of them revealed no significant relationships. The second clustering run of the study involved these 12 characters with numbers 1, 4, and 9 ordered with K = 1. Characters 8 and 11 were matrix types as defined in Fig. 2. The parameters established in each case were done so because it seemed biologically more accurate to have these states

Table 3.

Characters used in the second and final analysis of the study.

Character and Number	Character State and Number
1. Snout-vent length	1. Small (55-60mm) 2. Medium (61-67mm) 3. Large (68-73mm)
2. Left frontoparietal	1. Whole 2. Split
3. Right frontoparietal	1. Whole 2. Split
4. Gular patches	1. Males very broadly joined Females separate and reduced 2. Males broadly joined Females just separate 3. Males intermediately joined Females just or intermediately joined 4. Males just joined Females broadly joined
5. Ventral spotting	1. Heavy 2. Light 3. None
6. Abdominal patch shape	1. Square 2. Rounded 3. Shape indeterminable 4. Pigment lacking
7. Front profile	1. Crest high 2. Crest medium 3. Crest low
8. Femoral pores	1. 31 to 33 2. 34 to 35 3. 36 to 37 4. 38 to 39 5. 40+
9. Median dorsal scales	1. 43 or below 2. 44 to 45 3. 46 to 47 4. 48 to 49 5. 50 to 51
10. Lip color	1. Plain 2. Red or yellow
11. Circumabdominal scales	1. 42 or below 2. 43 to 44 3. 45 to 46 4. 47 to 48 5. 49 to 50 6. 51 to 52

Table 3. (cont .)

12. Scales between pores 1. 4
2. 5
3. 6
4. 7

Table 4.

Conditional probability table for right fronto-parietal (I) vs. location (J).

	I(1)	I(2)
J(1)	.50	.50
J(2)	.50	.50
J(3)	.50	.50
J(4)	.20	.80

Table 5.

Conditional probability table for location (I) vs. front profile (J).

	J(1)	J(2)	J(3)
I(1)	.80	.20	.00
I(2)	.60	.30	.10
I(3)	.10	.80	.10
I(4)	.10	.30	.60

Table 6.

Conditional probability table for abdominal patch shape (I) vs. location (J).

	I(1)	I(2)	I(3)	I(4)
J(1)	.90	.00	.10	.00

Table 6. (cont.)

J(2)	.80	.00	.00	.20
J(3)	.50	.30	.00	.20
J(4)	.00	.80	.00	.20

Table 7.

Conditional probability table for ventral spotting (I) vs. location (J).

	I(1)	I(2)	I(3)
J(1)	.70	.30	.00
J(2)	.70	.30	.00
J(3)	.50	.50	.00
J(4)	.20	.60	.20

Table 8.

Conditonal probability table for ventral spotting (I) vs. abdominal patch shape.

	I(1)	I(2)	I(3)
J(1)	.77	.23	.00
J(2)	.27	.54	..18
J(3)	.00	1.00	.00
J(4)	.17	.83	.00

interrelated in varying degrees (discussed in more detail in the next section).

Graph 1 is the result of the first clustering (November 25) and Graph 2 gives the skyline results of the December 19 clustering run. The Nov. 25 run had 3 levels

Fig. 2. The parameter assignments given to the matrix type characters 8 and 11.

Character 8

	1	2	3	4	5
1					
2	.40	.			
3	.30	.40			
4	.10	.20	.30		
5	.00	.10	.10	.30	

Character 11

	1	2	3	4	5	6
1						
2	.30					
3	.20	.20				
4	.10	.20	.40			
5	.10	.20	.30	.40		
6	.00	.05	.10	.15	.20	

before all the lizard were in the same cluster (C = .571). The clusters were subgraphed to illustrate all the connections between objects (see the simplified example in Fig. 3). The nodal distance arrays were analyzed and the specimens from Baca County were found to be more similar to one another than individuals of other populations were to each other (Table 9).

Table 9.

The percentage of the 5 most similar lizards which came from the same population is given for each location.

Baca County	66.7%
Fremont County	21.0%
Boulder Co., south	26.0%
Boulder Co., north	22.0%

The Dec. 19 clustering run had 20 levels which were also subgraphed. The subgraph for level 14 is given in

Fig. 3. A modified subgraph of level 2 of the Nov. 25 clustering run. Circled groups have all the members interconnected. Red lines connect specimens from the same geographical location; blue lines connect specimens from neighboring geographical locations if the two Boulder Co. populations are considered to be the same.

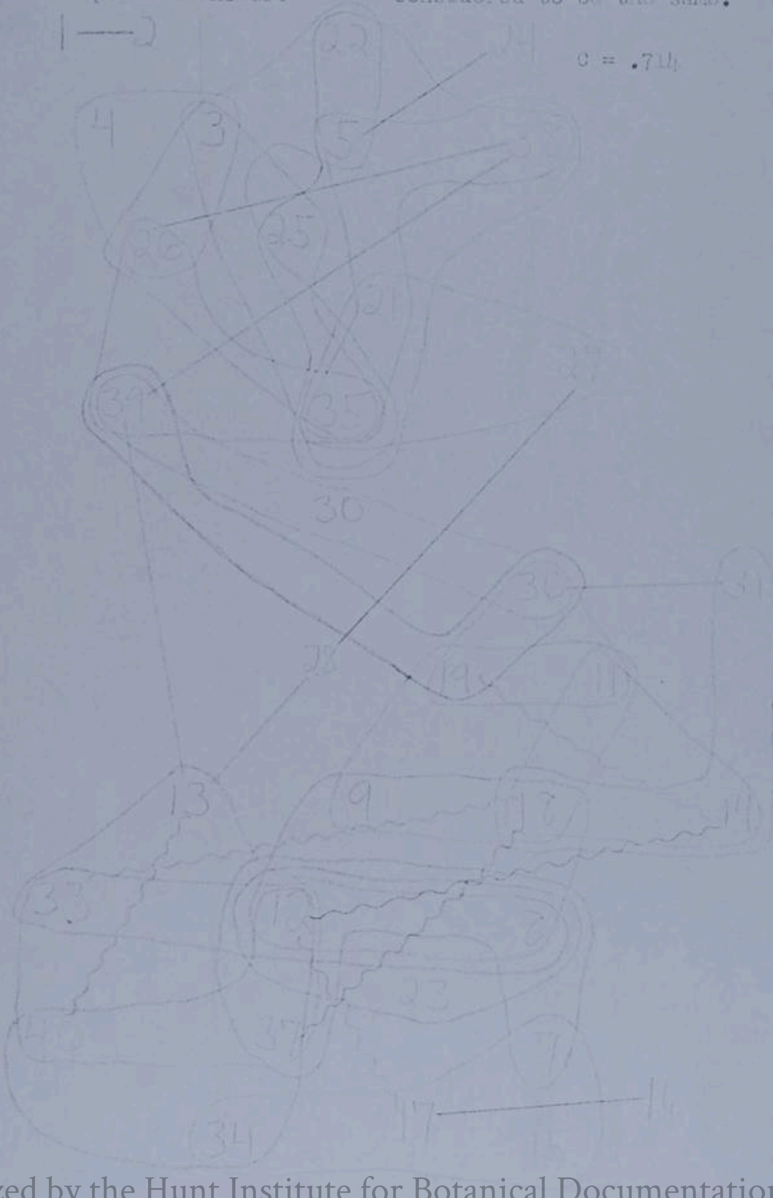


Fig. 4. A modified subgraph of level 14 of the Dec. 19 clustering run. Circled groups have all the members interconnected. Red lines connect specimens from the same geographical location; blue lines connect specimens from neighboring geographical locations if the two Boulder Co. populations are considered to be the same.

C = .700

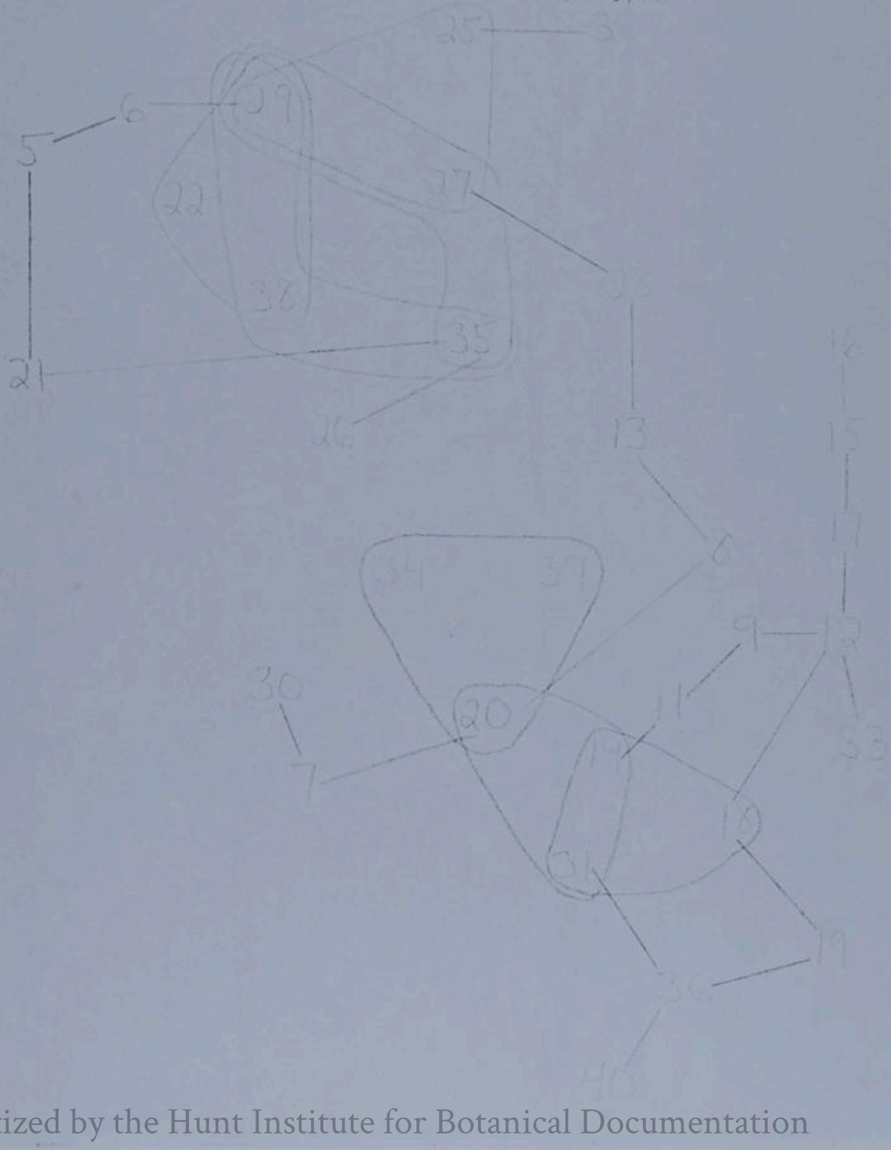


Fig. 4. Nodal distance arrays were not provided by the computer in this print-out. The C-value where all specimens joined into one cluster was 0.625, while that of the first run was 0.571 (see Graphs 1 and 2). The C-values where the first specimens joined were nearly identical in both runs.

Examination of characters separating and joining certain pairs of lizards revealed no significant patterns. The overall connections and clusters were not strikingly different between the two runs. All of the results obtained from the computer can not be presented here, but the data that is given is that which is significant to the following discussion.

DISCUSSION

The *undulatus* group of *Sceloporus* spp. is highly variable in form and is not easily characterized (Smith, 1946 and Stebbins, 1954). Structurally, the subspecies of *S. undulatus* can be separated only by average differences. Color and pattern are better used to separate the subspecies. Within the subspecies *S. u. erythrocheilus* there is also much variation in both structural and color pattern characteristics.

In selecting characters for this analysis an effort was made to obtain more than just the standard characters used to distinguish subspecies (characters 1, 8, 9, 11, and 12 in Table 3). Differences in color pattern were often difficult to use because of problems in expressing them and loss of information in processing and preservation. Chromatic sexual dimorphism is well developed in *Sceloporus undulatus erythrocheilus*. This was exemplified in this study by the character analysis of gular patches being more often joined in the males and separate in the females. The study was an attempt to classify irrespective of sex, so sexual dimorphism was carefully avoided in characters.

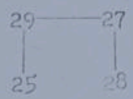
In most cases the designation of character states came easily. Several characters, however, should be discussed specifically. The states of character 1 (Table 3), snout-vent length, were obtained by graphing the distribution of lengths and selecting the clusters resulting. This character was ordered in the second run. Character 4 was also ordered because the states which had been adjusted for sexual dimorphism did not seem to be biologically distinct. Femoral pores (character 8) was established as a matrix type because some states appear to be much less related to neighboring states than others, especially at the extremes of the scale. This condition also held for character 11. The remaining characters were simple, excepting 9, which was ordered.

The results of character analysis did not reveal any great interdependency of characters which might suggest the existence of morphological types independent of geographical influence. This was upheld by the clustering results discussed below. Table 8 illustrates the best correlation that was found between morphological characters. The character analysis did, however, hint at a possible clustering by geographical location by a north-south cline.

The conditional probability tables given in Tables 4 through 7 (pp. 9 and 10) indicate the relation of certain characters to location. In each case three observations are made: 1). The two populations from Boulder County have extremely similar figures, 2). The Baca County population is close to having the opposite character state of the Boulder populations, i. e. the state with the highest probability in one population is that with the lowest in the other, and 3). The Fremont County population has its probabilities somewhere between the two above groups, always slightly more like that Boulder populations. These observations put together form a simple cline of change along a north-south line.

The Boulder populations are separated by Boulder Creek and might be suspected to be nearly identical. Personal observations have found lizards on both banks of the creek, suggesting possible gene flow between the populations. Several bridges across the creek and the tendency of young lizards to seek out new favorable rock habitats should be mentioned in support of this hypothesis. Indeed, the similarity of these populations has been varified in the results of the study. The remaining populations are more geographically separated as well as more morphologically distinct. They form the second and third steps in the cline. The reason that the Fremont group is closer to the Boulder group may be the apparent similarity in environments as opposed to Baca County which is to the east. The environmental factors involved may prove interesting for further research.

The similarity-graph-clustering technique resulted in no distinct groups distinguishable as morphological types in identical character states. By this it is meant that there is great variation as to which characters are similar among individuals in any one cluster. For example, at level 2 of the Dec. 19 run the following cluster forms:



The characters differing with each pair are as follows:

25 and 29 differ in characters 9 and 10
29 and 27 differ in characters 3 and 4
27 and 28 differ in characters 2 and 6

This limited example is typical of the study in that extensive interconnections did not occur in most clusters until C-value was about 0.70 and even then relationships were based on a full variety of characters. Therefore, no one character stands out as being exceptionally influential as working for or against a cluster.

The suggestion that geographical clustering occurs in this study can be investigated by analyzing the subgraphs and skylines. The first organisms to cluster in the Nov. 25 run did so at 0.856 and with a large moat (see Graph 1, p. 12). The small cluster had 2 from Baca and 1 from Fremont. The middle cluster had 4 from Baca, 1 from Fremont, and 2 from the combined Boulder group. The large cluster had 3 from Fremont and 12 from Boulder. This suggests a geographical pattern for clustering. Graph 2 also reveals this pattern, but in a more subtle way. The clusters here form much more slowly (i. e. fewer specimens join at each level) and the moats are less. The similarities between these two runs will become clearer as the subgraphs are discussed.

Fig. 3 illustrates the connections between specimens at level 2 of the Nov. 25 run. The subgraph has been modified to point out geographical relations with the Boulder populations taken together. Two major clusters stand out, one at the top and one at the bottom of the figure. All but two of the groups circled in the top cluster have more than neighboring members (therefore circled in gray). However, 7 of the 12 members of the entire cluster are from Baca County. The remaining members are 2 from Fremont (26 and 27) and three from Boulder (35, 38, and 39). It must be remembered that Boulder has twice as many representatives, so the overall picture is strongly Baca in nature. Specimen 39 is actually part of the articulation between the two major clusters; it is connected only to Fremont and Boulder specimens in the top cluster. Specimens 30 and 28 complete this articulation.

The bottom cluster of Fig. 3 had 15 members from Boulder, 3 from Fremont, and 1 from Baca (23). This cluster is almost entirely composed of specimens from Boulder and so the relations suspected from the character analysis are upheld. There is a good separation of Boulder and Baca, with Fremont as an intermediate. Now the results from the Dec. 19 run will be compared.

Fig. 4 is a modified subgraph of level 14 of the Dec. 19 run as C = 0.700, as compared to C = 0.714 in the above example. The same overall observation can be made here, although some of the specimens and connections have changed. The top cluster has 6 Baca, 3 Fremont, and 2 Boulder, while the bottom cluster has 16 Boulder, 2 Fremont, and none from Baca. Specimens 28, 13, and 8 form the articulation between the two groups. In conclustion, the results of the clustering reveal a grouping trend of specimens from the same or nearby geographical locations. There seems to be a north-south line of change in the relationships among specimens.

Exceptions to the above trend, such as 35 and 39, appear to have no regular relationship to the others in the group. It seems that due to the variability of characters in this subspecies, specimens like these will inevitably be chosen. If the purpose is to separate specimens into populations, better clusters may be obtained by using more characters and eliminating those which are more variable within a population. A classification of populations cannot be created easily or completely with these results; however, further study with more specimens and characters could result in separating even the Boulder populations to some degree.

SUMMARY

Sceloporus undulatus erythrocheilus was found to be somewhat taxonomically distinct as populations depending upon their geographical location. Populations separated by large distances (Baca, Fremont, Boulder) are better separated into clusters than populations closer to one another (the two Boulder populations). A cline occurs in the overall specimen relationship along a north-south line, although the variability of the subspecies causes some exceptions. This study suggests that further research could supply a more distinct classification of these lizards into population clusters.

Good paper – with some alterations, and with some consultation with Dr. Marlin on the biological interpretations, this could be published – I hope you will.
DR.

LITERATURE CITED

Estabrook, G. F. 1966. A mathematical model in graph theory for biological classification. J. Theoret. Biol. 12: 297-310.

______. 1967. An information theory model for character analysis. Taxon 16: 86-97.

______, and David J. Rogers. 1966. A general method of taxonomic description for a computed similarity measure. Bioscience 16(11): 789-793.

Maslin, T. Paul. 1956. Sceloporus undulatus erythrocheilus ssp. nov. (Reptilia, Iguanidae), from Colorado. Herpetologica 12: 291-294.

Smith, Hobart M. 1946. Handbook of lizards. Ithaca (New York), Comstock Publishing Associates.

Stebbins, R. C. 1954. Amphibians and reptiles of western North America. New York, McGraw-Hill Book Co., Inc.

Wirth, Michael, George F. Estabrook, and David J. Rogers. 1966. A graph theory model for systematic biology, with an example for the Oncidiinae (Orchidaceae). Systematic Zool. 15(1): 59-69.

TAXONOMY OF HOMINOIDEA -

a preliminary analysis

William H Reid

January 23, 1969
Biology 531

TAXIMETRICS

INTRODUCTION

The problem of Man's relation to other living primates and to fossil primate forms is one of both intense interest and contradiction: great efforts have been made attempting to determine the sequence of development which has yielded Homo Sapiens, but complete and viable system of classification exists.

It can be argued that an evolutionary sequence can be accurately be determined without recourse to formal Taxonomy and a system of classification, but the writer asserts that any such work must be suspect. Without submitting to the rigors of detailed specification of characters and the determination of valid taxa from these characters, there is no quantitative basis for determining the place of new specimens, no method for determining if conflicts exist, and no good way to approach the difficult question of differentiating Hominid species. Thus, an evolutionary study should include hierarchical classification of specimens involved.

The purpose of this study was to develop a classification for a limited number of fossil crania (crania being the most durable body part), and to check this against the historical evidence; i.e., are similar specimens or taxa closely related in time and (to a lesser extent) in location?

THE SPECIES PROBLEM

The usual definition of a biological species includes a statement that the species is a breeding population with a common, exclusive gene pool. There are a number of obstacles to using this as a criterion for primate species definition:

1) Much of the material consists of bones and fossil bones.

2) Many primates form bands (most particularly Baboons) and there is little breeding outside of the band. This is, in part, a result of the territorial behahior of the band. But the occasional movements of ♂ around

3) Most pretechnical human societies consist of relatively small groups or tribes with endogamous mating habits. These distinctions have nothing to do with organismic biology - as can be observed in most industrial or slave holding societies: breeding between previously distinct forms is common.

4) Laboratory breeding experiments with hominids are of questionable value, difficult to perform and would probably be a source of hilarity in the literature.

5) The question of time enters also: there is no evidence to suggest that, biologically, a modern man could not breed with a Cromagnon woman in one were available. There is a feeling that, because of 30,000 years (and vastly different cultures influence classification), that they do form different taxa - with the question of species put aside.

The tendency of Hominids to form endogamous, exclusive groups yields one very useful datum for those using fossil or archeological evidence: bones found in close association in undisturbed soil almost certainly belong to the same taxon.

No attempt to resolve the question of species will be made in this paper; the individuals studied will be assigned to a hierarchy of taxa.

THE OSTEOLOGICAL MATERIAL

General Considerations-
Any study utilizing fossil material must rely heavily on the bones and the teeth. Those that are most frequently preserverd and most useful to study are:

- the cranium,
- the pelvis,
- the long bones of the leg.

This study has been confined to the cranium since that is most frequently preserved and most commonly studied.

In using the cranial bones it is important to keep in mind that they reflect the organism's genetics less accurately than many other parts of the body. The points given below were drawn from Torrey(1967):

The bones of the brain case ossify after the brain and optic capsules have formed, and reflect the structure of these organs.

The mandible and maxilla are derived from the first visceral arch of the embryo and may be determined by different genetic factors than the brain and brain case bones.

The teeth are derived from inpocketings of the ectoderm and may well be determined by different genetic factors (for example, the problems of wisdom teeth which dont fit in the jaw of modern man).

Bones are strongly influenced by environmental and cultural factors; e.g., diet and artificial deformation.

Bones are influenced by muscle attachment. For example, the ridges on a dog's skull occur at the attach lines of chewing muscles. If these muscles are detached at birth, the ridges do not form.

From this it can be seen that a number of different characters should be used, and that these, reflecting a number of influences, will not fall into well defined states.

Characters-

The characters used in this study were drawn from those used in physical anthropology for the general description of crania. They are as follows:

Canine Prominance - This character is most useful for distinguishing between very early hominids and the later forms.

Brow Ridge Development - The bony ridges above the eyes have become reduced only in the last 100,000 years of human development - much later than some of the other cranial ridges. In the writers opinion, it may reflect changing selective pressures with changing modes of combat.

Prognathism- The projection of the lower face and mandible has been steadily reduced with time among hominids, but still exists among Homo sapians; particularly some Africans.

Facial Angle - The angle of the upper part of the face shows wide variability among the taxa, but is determined by a number of different structures.

Saggital Crest - This crest is determined to some extent by eating habits and the chewing muscles. It disappeared with Sinanthropus to a large extent, but occurs among apes, Gorilla gorilla for example.

Gabling of Paretial Bones - Among H. sapiens this shows some variability, among more primitive forms it occurs, but is sometimes emphacised by the saggital crest.

Occipital Prominance - This has been steadily reduced with more erect posture and decreasing neck musculature.

Forehead Slope - The forehead slope reflects the relative importance of the diencephalon (cerebrum) in the cranial vault.

Opisthion-Inion Crest - This crest reflects neck musclature, but is often hard to detect in photos or damaged specimens.

Cranial Index - The width to length ratio is most useful for differentiating among undeformed, human crania.

Shape - difficulty was found in using this: it is hard to judge and is redundant with the cranial index.

Brow Ridge Form - The brow ridges are divided in some forms, and in H. sapians they are usually to indistinct to clearly decide.

Specimens -

Thirtyfour specimens were used. They represent as broad a selection as possible for which adequate photographs could be obtained. ~~Table~~ Figure 1 shows the locations and relative ages along with the summary data given below. No attempt will be made to describe the crania here; space and time do not permit.

#2 thru 8: Bambandyanalo, British East Africa, a paleolithic Bantu site, Broom(   ).

#9 thru 11: Sinanthropus pekinisis, Weidenreich (   ).
#12 ans 13: Paranthropus robustus (or Australopithicus robustus), Swartkrans, South Africa,

#14: Texepan Man, Texepan, Mexico, de Terra (   ).

#16: Pithecanthropus erectus, Java, Weidenreich(1940).

#17 thru 20: Solo Man, Solo River, Java, von Koenigswald(1951).

#22 Gazelle peninsula, New Britain, a relatively recent cranium, Mac Curdy (1914).

#23 thru 25: upper paleolithic crania, China, Angel (1966).

#26 thru 30: Paleoamerican crania from four sites in the United States, Angel(1966).

#31 and 32: "old grave" Eskimo crania from Labrador, Stewart(1939).

#33 thru 37: recent eskimo crania from Labrador, Stewart(1939).

RESULTS

The suggested classification is shown in figures 2 [illegible].

Note that two forms of the modern eskimo have been suggested. One of these forms is distinctly different, and might represent a hybrid with europeans.

CONCLUSIONS

A classification of <u>Hominoidea</u> which does not conflict with historical evidence is feasible. Figure 3.

More study is required to develop adequate characters.

A number of characters should be used, and these should be adequate for a wide range of cranial geometries.

In selecting characters, reference should be made to human morphogenisis so that they will reflect as wide a range of genetic influences as possible.

Figure 1
Locations and Relative Age of Specimens

Continental Area

North America	Southeast Asia	Continental Asia	Africa
Recent Eskimo			
Old Eskimo			
Paleoamerican	Gazelle		
Texepan		Paleolithic Chinese	Bambandyalo
	Solo Man	Sinanthropus	
	Pithecanthropus erectus		
			Paranthropus robustus

Relative Age

↓

Increasing

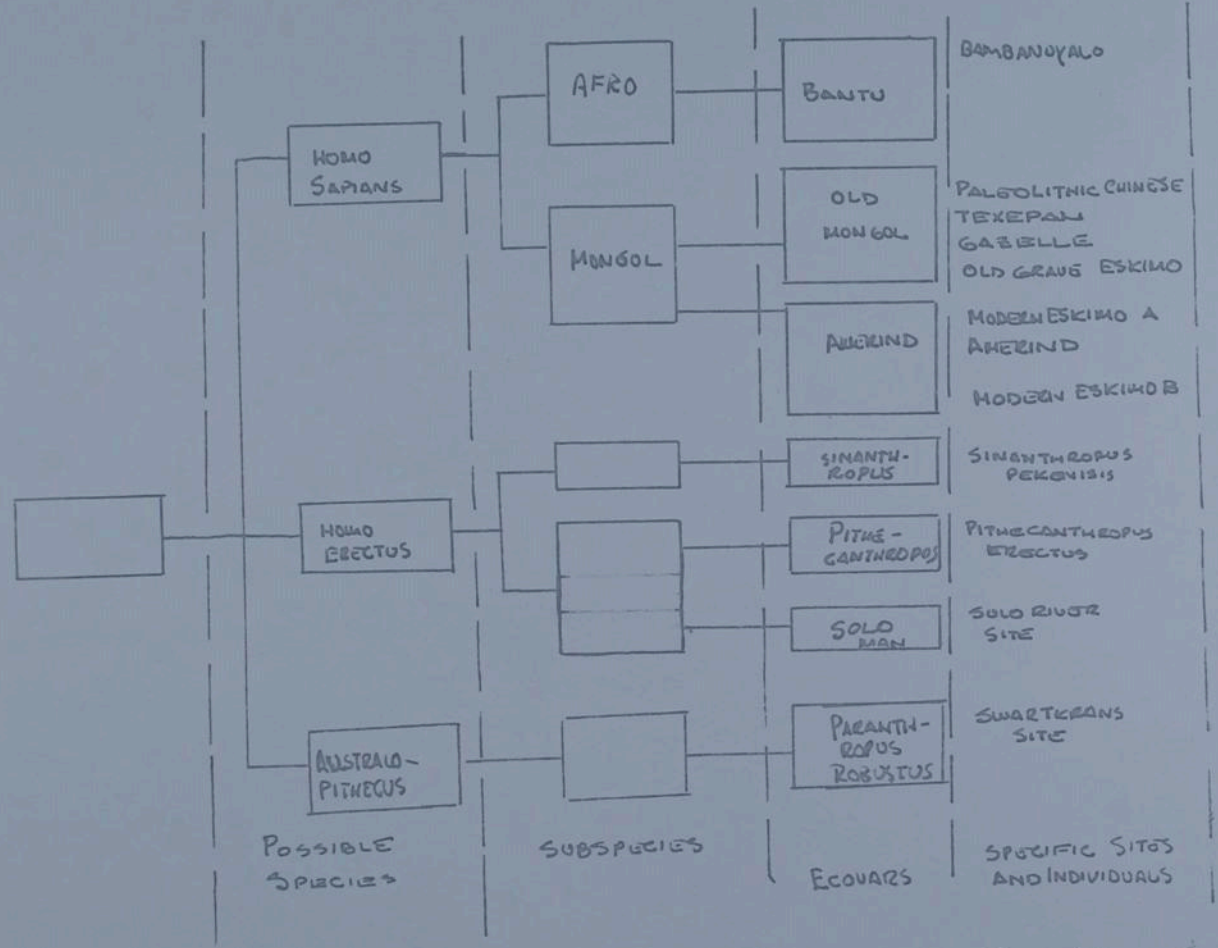
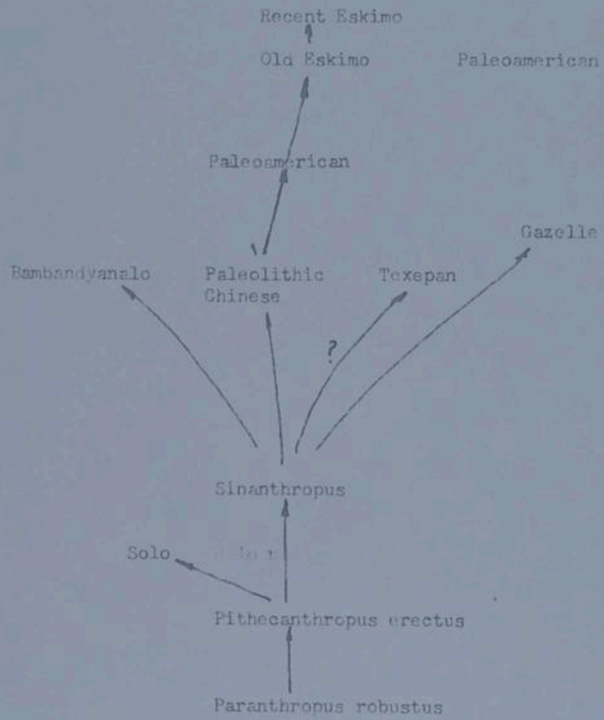


FIGURE 2 HIERARCHICAL CLASSIFICATION FOR STUDY SPECIMENS

Figure 3

An Evolutionary Sequence Permitted
by the taxonomic data

Note: This sequence is limited to the forms included in the study, and was developed by relating forms with the greatest taxonomic similarity. The sequence shown is not the only one possible.



A STUDY OF THE VALIDITY OF THE VARIETIES OF

PEDIOCACTUS SIMPSONII (ENGELM.) BR. & R.

By GERALD ARP

INTRODUCTION

Although the Cacti have been repeatedly studied by numerous investigators, a quick scanning of the literature will indicate vast amounts of confusion arising from all levels of their classification. There are fundamental reasons for this confusion.

(1) The basic structures of these plants are quite simple and several of them are continually repeated (i.e. ribs and areoles). Thus, a minute, barely perceptible physical change in one of these organs will cause tremendous alterations in the overall appearance of the plant.

(2) The plants frequent areas of drastic climatic variability causing a single genetic group to be differentially influenced by a number of contrasting conditions within a small area with the resultant array of phenotypic variabilities.

(3) The Cactus Family contains several very large genera which are still evolving very quickly (i). Hence, the members are able to draw their genes from a very large "genetic pool" creating very plastic species from which diverse individuals may develop.

(4) Basic flower types throughout the family vary but little (ii) and within a genus they may not change at all. Therefore, the floral characters are less useful in the classification of the species.

A confused polymorphic group results from the above, which defies the traditional means of classification. Add to the problem a lack of usable herbarium materal from which studies are usually based and the constant erection of new taxa upon single characters or single plants and chaos results (iii).

It is this highly confused nomenclature which initiated my interest in the varieties of Pediocactus simpsonii (Engel) Br. and Rose (iv). My choice was based upon these criteria.

(1) Pediocactus simpsonii is essentially monotypic though new discoveries along the Colorado - New Mexico border are now credited to this genus (v).

(2) The plants are easily available as they are native to most mountain areas of Colorado and are also well represented by several local collectors in Denver whose gardens contain almost every conceivable form of the plant.

(3) Three excellent varieties are present, each of which is easily identified. The first is var. simpsonii which constitutes the variety upon which Engelmann made the first descriptions, the second is var. minor, a high mountain variety, and the third is var. robustior, vigorous plant from Oregon (vi). The last variety is so distinctive that one could almost consider it another species.

From this study I hoped to gain some idea of what constituted a variety and what might be some of its limitations. Upon assembly of the plants, work began. The first phase of the study concerned itself with an analysis of the validity of the various commonly used discriptors. Using the Character Analysis computer program, I was able to discover several lesser used discriptors that are highly diagnostic and several commonly used ones that have little use in the diagnosing of the assorted varieties. From this I developed two types of discriptors. The first group is vital to the keying of the varieties and, thus, constitutes a key (see key below), and the second type is useful largely in the variety description. The second portion of the study was an attempt to determine how coherent each variety is. It is from this second portion of the study that I hoped to determine the validity of the three Pediocactus varieties.

RESULTS AND DISCUSSION

From Character Analysis I was able to develop an extremely simple key which accurately divides the species into its three component varieties. The following key is based upon certain "keying characters" which were found by the Character Analysis process to be very useful in segregating the varieties. To be sure, the key is simple, but it professes only to indicate the identity of the plants of the study.

a. Plants with ribs or coalesced tubercules -

Pediocactus simpsonii
var. *robustior*

aa. Plants without ribs - b.

b. Central spines very slender and beet red -

Pediocactus simpsonii
var. *minor*

bb. Plant body completely obscured so that no body features are clearly visible, also none of the above features are in evidence -

Pediocactus simpsonii
var. *simpsonii*

All of the rest of the discriptors of the study help to describe the varieties (description characters). Descriptions of the three varieties are as follows:

Variety robustior - Single to clustering, globose, ribbed, plant body not obscured by the dull black central spines, and native to Oregon and Washington. There exists an intermediate form in Idaho, which while less robust still maintains the ribs and black spines.

Variety minor - Single to clustering, compressed to short globose, tuberculate with shiny beet red spines which superficially appear black and do not obscure the plant body. A very inconspicuous plant blending well with the black soils of the high mountains of central Colorado.

Variety simpsonii- Single to clustering, globose to short cylindric, and very spiny. The centrals spines may be white, yellow, brown, cinnamon or any color in between. The plant body is completely obscured by the spines. This variety is the most commonly found, occurring in Colorado, Wyoming and Utah from the plains to the montane zones.

The second part of the project concerned itself with a study of the validity of the variety classification and an attempt to place several unusual forms of this species in one of the variety classes, if possible. From the graph cluster analysis I obtained three nice clusters which accurately segregated the varieties. The seedlings and transitional forms of the varieties did not fair so well. As feared, they tended to move out of the well defined variety classes and into a zone intermediate between two varieties. Thus, it was apparent from the first that the mature characters of each plant were governed more by the climate (phenotypic response) than by their genes (genotypic response). These sorts of problems were especially evident in the various forms of var. simpsonii. There was, however, a case in which a var. minor plant appeared in a var. robustior cluster, and another in which several Idaho plants occurred in both var. minor and var. robustior clusters. I interpret this as meaning that var. robustior evolved from var. minor as the var. robustior moved to the Pacific Northwest into more temperate climates. I believe that the direction is away from Colorado instead of toward it because more forms of Pediocactus occur in Colorado, the population is greater in a less favorable area, and the Colorado plants lack the more advanced ribs and heavy spines of the var. robustior.

The plants of central Idaho appear intermediate between var. minor and var. robustior, but this may be due to their age (all were seedlings or young mature plants). They have been included with var. robustior by virtue of their ribs and slender black central spines. In the beginning I felt that var. robustior might be a sub-species due to its characteristic appearance. Research does not support this, as there is a clinal change from var. minor to var. robustior. I also felt that var. minor might just be a "form", again studies offered no support. ~~As the map shows~~, There are definite regions where var. minor exists and in these regions var. simpsonii is not to be found. This, along with its distinctive appearance, tend to reinforce the validity of its status as a variety. The var. simpsonii represents some of the most interesting studies. Within its range several forms exist, including the "cinnamon spined" forms and another in which all spines are compressed creating a harmless ball. Both forms are sporadic in normal populations and do not segregate into separate populations. Therefore, they could not be viewed as anything more than "forms". Included, also, in the study were four very dwarf plants whose affinities were not known. Two of these plants grew in Salida, Colorado, on the eastern slope and two were from Paradox, Colorado, on the western slope. The two from Paradox are members of a population so peculiar that various authorities have felt they might represent the diminitive and elusive Pediocactus knowltonii of Northern New Mexico (vii). Thanks to the cluster analysis, enough bits of information were processed that the true identities are now revealed. Both represent the var. simpsonii. This last example indicates the worth of this new system of Taxonomy, as without it no positive identification could be made of these retarded juvenile plants.

Many seedlings were included in the study. These seedlings are difficult to classify for they have few mature features. Again, the computer methods helped place the plants in their proper places and, also, used them to tie together many of the forms that make up var. simpsonii.

CONCLUSIONS

From this study I conclude that there are indeed three varieties in the species Pediocactus simpsonii, and they are:

1. Pediocactus simpsonii var. simpsonii Benson.
2. Pediocactus simpsonii var. minor Engel.
3. Pediocactus simpsonii var. robustior Coulter (viii).

I may also conclude that these varieties are sound in the biological sense of the term. I feel the var. robustior has been derived from var. minor. Lastly, I believe none of the conclusions drawn here would have been obtainable in the amount of time available without the tremendous assistance given by the computer through the Taxometrics methods used.

CRITICAL COMMENTS

Without the Character Analysis and Cluster Analysis systems I doubt that this study would have been worthwhile. Thanks to these methods I was able to consider vastly more material than I ever could hope to, using standard methods. From these considerations, I was able to turn several "novel" facts into valuable criteria for classification. Further, I was able to "see" the evolution of the vigorous var. robustior from the conservative var. minor. The conclusions drawn are, of course, tentative and apply only to the plants of the study. However, through judicious selection of the members of the study I feel I was able to make valid appraisals of the plants which represent this species. In order to make the conclusions, indicated here, scientifically valid and the key accurate for all plants of the species, another study of expanded scale will be necessary.

ACKNOWLEDGMENTS

I would like to thank all the members of the Taxometrics Lab. for their unfailing interest in my project and Mrs. Mary Ann Heacock of Denver, whose assistance in obtaining plants and critical locational data made this study possible.

LITERATURE CITED

i. Dawson, E. Yale; The Cacti, Wm. C. Brown Co., Publishers; Iowa 1963 page 8.

ii. Buxbaum, Dr. Franz; Morphology of the Cacti, Section II, The Flower, Abby Garden Press, Publishers; Pasadena, Calif., 1950.

iii. Backeberg; Kakteenlexikon, 1965. This book represents Dr. Backeberg's classification.

iv. Britton, N. L. and Rose, J. N., The Cactaceae, Dover Press, Publishers; N. Y., N.Y. Reprint 1963, page 90.

v. Benson, Dr. Lyman; A Revision and Amplification of Pediocactus I, Cactus and Succulent Journal XXXIII no. 2, 1961, page 49.

vi. ibid.

vii. ibid.

viii. ibid.

1-16-69

Course A
Paper A
Talk A-

A RESEARCH PROBLEM IN BEHAVORIAL TAXONOMY

done at

COLORADO UNIVERSITY

during

SPRING SEMESTER 1968

by

GARY D. CAMPBELL

with the direction of

DR. D. J. ROGERS
and the Taximetrics staff

Gary -

Good paper - but I would want a little more discussion in the introductory remarks which said something about:

1) Why bother with a classification scheme in psych.?

2) Can any generalities be developed in the psychological milieu after some classificatory work has been done?

Then in the conclusions, or summary, or other heading at the end of the paper

1) try to come back to the introductory generalities -

2) Do a little "wild blue yonder" speculating.

A RESEARCH PROBLEM IN BEHAVORIAL TAXONOMY

INTRODUCTION

The problem that I have studied this semester was a preliminary investigation. A special research technique was used to develop a classification system for job applicants. The actual materials with which I worked were Personal History forms completed by the applicant. Designed by Worthington, Hurst & Assoc. they reflected each applicant's encounter with the unstructured design of the form. This allowed him to express both relevant facts about himself as well as hints about his personality.

? Poor ? construction

Population: 50 subject's PH forms were selected. There were ten from each of five groups: The accounting field, engineering, hospital supplies salesmen, a random group and finally ten Redfield insurance salesmen (aggressive, hungry types).

HYPOTHESES:

1. The best basis of a classification are characteristics which correlate and interrelate with one another and not descriptors which have no relation to one another.

2. The individual characteristics which form the basis of the classification are actual observables and the groups into which persons are to be classified are typified by a more abstract definition.

3. A mathematical computing model can be used to compare characters selected to describe the various aspects of the PH form. A computing model can also be used to perform the classification and form the objects into groups using this set of characters for an input.

4. The characters one may derrive from the objects to be studied have various properties which will require a model that can make comparesons between characters of the following types:

a. Nominal Characters which have actual states that are exclusive, such as the way in which a name may be written, i.e. John Jones, John J. Jones, J. John Jones etc.

b. Ordinal Characters, i.e. Age at marriage; 21, 22, 23, 24, etc.

c. Matrix Characters, i.e. partial similarity exists between some different states but not all of them and perhaps in varying amounts.

There are two approaches to this type of problem which will be used to complement each other. These are Character Analysis CHARANAL and Cluster Analysis or GRAPH. They have been developed originally to help in biological classification by Dr. D. J. Rogers and his staff at Colorado University, and recently have been utilized by researchers in other fields. This problem is the first attempt to employ it in the field of Psychology. CHARANAL and GRAPH are computer programs designed to assist a researcher in forming a classification system. Through an iterative process involving three phases, the researcher works with the computer to discover a working classification system. These three phases are: 1) the construction of characters and character states, 2) Character Analysis and 3) Cluster Analysis.

Initially the researcher has a vague notion of what he wants to gain from a system of classification. In this case it is a person[al] evaluation of the individual represented by his PH form. This means that our final product should be a method to systematically describe the PH form in such a way that a cluster analysis of a number of these forms places individuals together with respect to similar overall characteristics and separates those which are unlike in this respect. To discover the best system of PH form description, one begins by advancing a test system for a group of objects. He then submits this to the computer for CHARANAL. These results are examined and he makes changes which take advantage of relevant information and dispose of randomizing information. He re-runs CHARANAL and repeats the process. When he thinks he has a fairly good system, he submits his data to a GRAPH run for a Cluster Analysis. From this he examines the clusters to gain clues about the way in which his system actually groups the individuals. He may use the results of the Cluster Analysis to modify his data input even further and go back to another Character Analysis. Eventually the results of his work can be checked (by computer with these same programs if necessary) against direct evaluations of the individuals represented in the study and in this way a valid working system developed.

CHARANAL or CHARACTER ANALYSIS

When a classification system is designed to group objects so that those which are similar to more abstract criteria than can actually or practically be observed it must accept as its basis some set of descriptors for each object. Although the variation between objects dictates the actual relevant descriptors for each, the researcher must first define his general case. He must select a set of observations which can be made about each object. These observations can be measurements or the determination of an actual state of a particular aspect of the object. The aspect being observed or measured is called a character, the various forms that it can take among the objects in the study are known as character states. When a character has either ordinal or cardinal properties these properties may be maintained to an extent by assigning a sufficient number of character states. This will, however, cause a greater splitting if each state remains distinct. To counteract the splitting that occurs as more states are used to define a scalar measurement, the researcher defines for this type of character the number of neighboring states which in his opinion bear a similarity to each state. This relation drops off linearly. If twenty-five states are used to describe a linear measurement, he might decide that states more than five apart are completely distinct, but as they get closer than this they become more and more similar.

Another case which can occur is that where a character does have distinct states, but some of them bear a partial similarity to each other. When this type of character is encountered, the researcher can construct a matrix and define for each possible pair of states the similarity he feels they have. With these facilities, the program can make comparesons between any of these types of characters.

Character Analysis, then, is a computer program which compares pairs of characters by examining all the objects described by the two characters. It automatically computes the following:

1. The amount of information contained in each character.
2. The probabilities of the individual character states.
3. The conditional probability of the states of one character given a particular state of another character.
4. The information contained in a character given objects in a particular state of another character.
5. The information contained in common by two characters.
6. The "distance" between two characters which varies from 1 to 0 as the character pair goes from measuring completely independant aspects to the point of measuring completely identical aspects.

By studying these things, the researcher gains a knowledge of what the characters he has selected for description are doing with relation to one another. By noting the fact that certain states of several characters show a relation he can better tell if they are describing the intrinsic or more abstract concepts which he wants his classification to be based upon. When he feels that the relevant concepts are being represented and that a complete enough description is contained (but no more than necessary) in his set of characters, he is ready for the next step.

GRAPH or CLUSTER ANALYSIS

If the description which has been developed is to provide the basis for a classification then objects which are similar to the intuitive or abstract criteria must be, in fact, similar when described by the set of characters thus far developed.

GRAPH is a computer program which has proven to be a valid tool for research problems of this type. Cards which contain the information describing each object form the data input. Parameter cards are designed to designate the characters for which similarity between states is not 0 for unlike states, but is some value greater than 0 (which normally describes the similarity between two objects coded into a different state) and less than 1 (which always describes the similarity between two objects in the same state). This is done, as previously mentioned, by naming the character and giving the number of neighbor states which are to be counted as similar or by giving a matrix which contains the similarities.

The computer program computes the overall similarity between objects by averaging these individual character similarities. It computes a similarity for each pair of objects and orders these pairs from the highest similarity to the lowest. From this point, five types of output can be called for to aid the researcher. The first is simply the ordered character pairs and their similarities. The second is the ordered similarities and their character pairs. The third is a page by page print-out of the clusters. It is this print out which is the most useful in determining the grouping arrangements which follow from the object descriptions. The first page of this print-out shows the similarity value at which the first pair or pairs of objects is formed. It gives this value, names the pair or pairs and tells for each pair or group the difference between that similarity level and the similarity level at which the next objects joins the group by forming a pair with a member in the group. The researcher can, by this means, draw pictures depicting his objects and the connections between them formed by the existence of a high enough similarity. One such picture appears in this report showing objects clustered at a level of .667 or higher. The fourth output is a list of each object and the ten objects most similar to it with the similarity values they make when compared to it. The fifth type of output is done by a Calcomp Plotter and it draws a "Skyline" which shows groupings, the levels they enter the cluster and how the clusters grow and combine at progressively lower levels.

With these data, the researcher is in a position to see which of his objects group together and whether he has a definitive clustering scheme and thus a classification system. If things are grouping as he would like, he can check to see if they are grouping logically by referring back to their original descriptions. If the clustering is not what he wants this analysis will tell him what it actually is doing. Sometimes if the Graph run is really pre-mature most of the useful information will come from examining the various similarity listings.

On pages 4 and 5 is a list of the 15 characters which I selected from an original list of 26. I chose this group on the basis of how much information each character seemed to be sharing with the other characters. A preliminary CHARANAL run with all 26 characters provided this information and allowed me to number the 15 chosen for this report in the order of best information sharers first. Pages 6 and 7 are the condensed results from the CHARANAL printout (only for the first couple of characters). Pages 8 and 9 are some of the pictorial results of the GRAPH printout, which are hand drawn from numerical information.

K_1 Name 1.90 IU
1. John Jones .20
2. John J. Jones .52
3. J.John Jones .02
4. John Jones III .02
5. Jones, John J. .04
6. John Joseph Jones .18
7. John Jones Jr. .02

K_2 Group 2.20 IU
1. Accountant .20
2. Engineers .20
3. Salesmen for Ethicon Sutures .20
4. X (random) .20
5. Redfield Insurance Salesmen .20

K_3 Style 1.66 IU
1. Block throughout .60
2. Script " .08
3. Writing " .18
4. Changes on page 1 or 2 .02
5. Changes on page 3 or 4 .12

K_4 Salary Ambitions 1.85 IU
1. Increase 0% to 50% .12
2. " 50% to 110% .30
3. " 120% to within reason .48
4. Skyhigh .10

K_5 Extra Curricular Activities 2.03 IU
1. Athletic Body Contact .26
2. " Individual .20
3. Quiet Constructive .40
4. " Vicarious .08
5. None or other .06

K_6 Title 2.20 IU
1. Line or lines thru both inappropriate titles .42
2. Single lines or "X's" thru " " .20
3. Quick lines (upstrokes) " " " .02
4. Heavy blocking out of each " " .12
5. Any method of dealing with appropriate title .14
6. Title not dealt with .10

K_7 Age at first Marriage 1.97 IU
1. 19 - 20
2. 21 - 23
3. 24 - 27
4. 28 and over
5. can't tell

K_8 % letters to words in margin p.4
1. 1 - 9% 1.79 IU
2. 10-19%
3. 20% or more
4. none at all

K_9 Highest Grade 1.09 IU
1. High school or less
2. Some college
3. Grad or more

K_{10} Total words on page 2 1.87 IU
1. 20 - 40
2. 41 - 50
3. 51 - 60
4. 61 and over

K_{11} Tattoo marks .71 IU
1. Yes
2. No
3. Blank

K_{12} No. negative words 1.95 IU
1. 3 - 7
2. 8 - 10
3. 11 - 12
4. 13 & over

K_{13} WHA rating 1.46 IU
1. A
2. B
3. C
4. D

K_{14} Sibling Relation 2.28 IU
1. Only child
2. Youngest (w/sister)
3. Middle "
4. Oldest "
5. No sister

K_{15} Scars 1.28 IU
1. Unnecessary mention
2. No or necessary mention
3. Blank

DATA EXPLANATION

D1 CHARACTER PAIR (I, J)

Probability of I_x given J_y increases from 20% to 60%

$P(I_x / J_y)$ 20 → 60 — where 20% is the actual probability of I_x and 60% is conditional

x and y are states of I and J

Pair		
(1, 2)	$P(I_1/J_5)$ 20 → 70	$P(I_1/J_3$ or $J_4)$ 20 → 0
	$P(I_5/J_1)$ 4 → 20	$P(I_6/J_5)$ 18 → 0
(1, 3)	$P(I_2/J_2)$ 52 → 100	$P(J_2 / I_2)$ 8 → 15
(1, 4)	$P(I_1/J_1)$ 20 → 67	$P(I_1/J_4)$ 20 → 40
	$P(J_1/I_1)$ 12 → 40	$P(J_4/I_1)$ 10 → 20
(1, 5)	$P(I_1/J_5)$ 20 → 67	$P(I_6/J_4)$ 18 → 50
	$P(J_5/I_1)$ 6 → 20	$P(J_4/I_6)$ 8 → 22
(1, 6)	$P[(I_3)+(I_5)+(I_6)/J_4]$ 26 → 56	$P(I_1/J_1)$ 20 → 6
	$P(I_1/J_5)$ 20 → 43	$P(I_1/J_6)$ 20 → 60
(1, 7)	$P(I_1/J_5)$ 20 → 66	$P(J_1) = 8$ $P(J_2) = 20$ $P(J_3) = 44$ $P(J_4) = 22$ $P(J_5) =$ [illegible]
		But Given I_1 → 0 → 0 → 50 → 30 → 20
(1, 8)		$P(J_1) = 40$ $P(J_2) = 34$ $P(J_3) = 18$ $P(J_4) = 8$
		But Given I_6 → 22 → 33 → 44 → 0

(1, 11) TATTOOS vs. NAME

	$J_1 = 4$	$J_2 = 86$	$J_3 = 10$
I_1		70	30
I_2	8	84	8
I_3		100	
I_4		100	
I_5		100	
I_6		100	
I_7		100	

(1, 15) SCARS vs. NAME

	$J_1 = 34$	$J_2 = 58$	$J_3 = 8$
I_1	40	40	20
I_2	31	61	8
I_3	100		
I_4	100		
I_5		100	
I_6	22	78	
I_7	100		

PROBABILITY OF STATES OF CHARACTER I = 1 (NAME)

$I_1 = 20\%$ $I_2 = 52\%$ $I_3 = 2\%$ $I_4 = 2\%$ $I_5 = 4\%$ $I_6 = 18\%$ $I_7 = 2\%$

DATA

(2, 3) $P(J_1/I_2)$ 60 → 100 $P(J_3/I_1)$ 18 → 30

$P(I_1/J_5)$ 20 → 33 $P(I_3/J_5)$ 20 → 33

(2, 4) $P(I_5/J_1)$ 20 → 83 $P(J_1/I_2 \text{ or } I_3 \text{ or } I_4)$ 12 → 0

$P(J_2/I_2)$ 30 → 50 $P(J_3/I_3)$ 48 → 80

$P(J_4/I_3)$ 10 → 0

(4, 5) $P(J_5/I_1)$ 6 → 33 $P(J_2/I_4)$ 20 → 40

$P(I_1/J_5)$ 12 → 66

ALL NUMBERS BELOW REFER TO PEOPLE <u>NOT</u> %

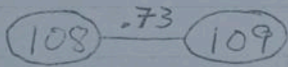
except for Subscripts and character Pair in Brackets

Group vs. Extra Curricular Activities

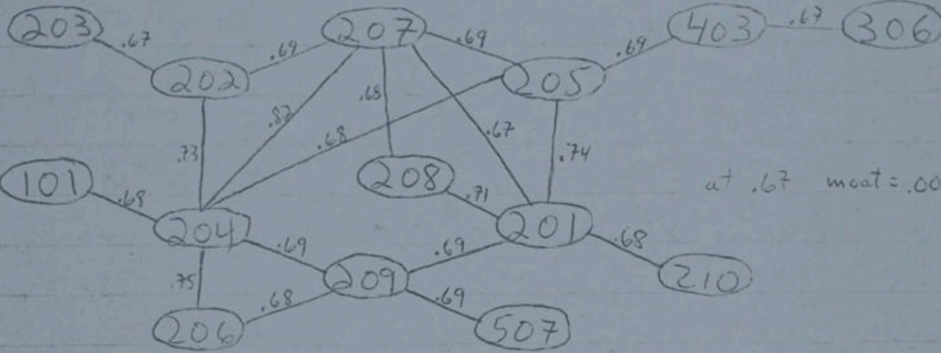
(2, 5)	$I_1 = 10$	$I_2 = 10$	$I_3 = 10$	$I_4 = 10$	$I_5 = 10$
$J_1 = 13$	1	0	5	3	4
$J_2 = 10$	3	2	0	3	2
$J_3 = 20$	4	8	4	2	2
$J_4 = 4$	2	0	1	1	0
$J_5 = 3$	0	0	0	1	3

STYLE vs. Extra Curricular Activity

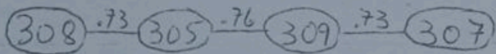
(3, 5)	$J_1 = 13$	$J_2 = 10$	$J_3 = 20$	$J_4 = 4$	$J_5 = 3$
$I_1 = 30$	4	8	14	3	1
$I_2 = 4$	2		2		
$I_3 = 9$	1	2	4		2
$I_4 = 1$	1				
$I_5 = 6$	5			1	



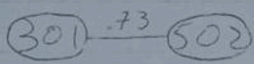
at .73 moat = .10



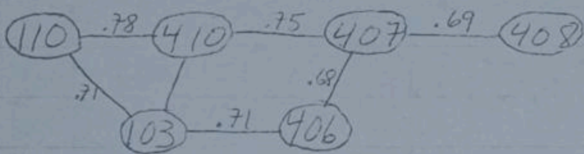
at .67 moat = .007



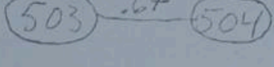
at .73 moat = .05



at .73 moat = .08

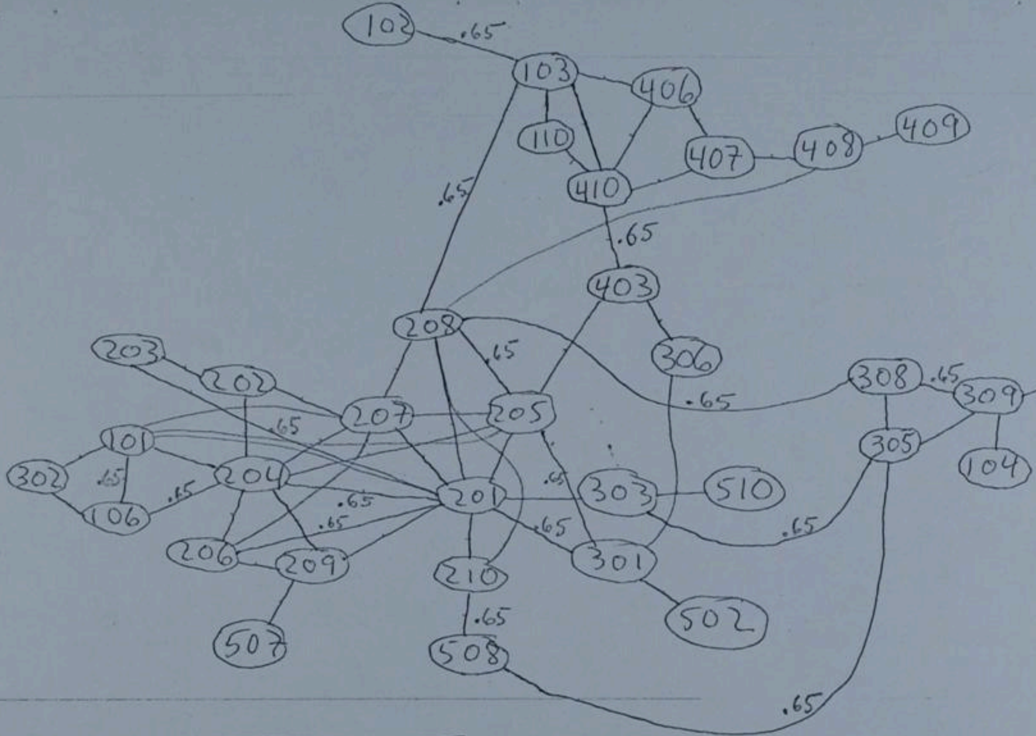


at .68 moat = .03



at .67 moat = .09

"moat" refers to the similarity difference between the level depicted and the level at which the next object attaches



MOAT = .02

-9-

LEVEL 20 (BREAKING POINT)

ALL UNMARKED CONNECTIONS ARE .66 OR HIGHER
SEE PREVIOUS LEVELS

RESULTS (DISCUSSION)

CHARANAL

The 15 characters which seemed to relate best with all the other characters in terms of information sharing were used in the Character Analysis. The motive for selection of information sharing characters is that these offer the best chance of speaking to the underlying more abstract description which should form the actual classification. This is true since none of the characters independently contain this description and cause and effect between characters is less likely than some set of outside factors influencing both characters of a pair in a like fashion.

The results listed on previous pages are items noted from the CHARANAL conditional probability matrices. I selected the pairs containing the first character and all of the rest and the first five paired with each other. Time did not permit me to make a more complete investigation, and at this point additional steps in the research process are indicated so that more conclusive results might be obtained. I will discuss these at the end of this section.

In the two pages of CHARANAL results, the most significant results of the program are listed. The first pair (1,2) is the Name vs. Group. Refer to the list of characters and character states on pages 4 and 5. The data is written in standard mathematical notation, giving the conditional probability of the state of character I (or which ever first appears) given only the objects that are in the indicated state of character J. For example, the first expression on the page reads: (1,2) $P(I_1/J_5)$ 20 → 70

This means Character I=1, which is the Name, is being compared to Character J=2, the Group. The probability of character 1, state 1 given character 2 state 5 goes from .20 to .70, in other words the initial probability of I_1 was 20% but given objects in J_5 it increases to 70%. A good example of the information available through this technique can be gained from evaluating these results yourself and seeing the kinds of things which are pointed out.

It can be seen that certain states of each of the characters follow the general distributions very closely (those are all of the ones I did not list) while others (the best of which I did list) deviate drastically. In those cases where selection of people who are in a particular state of one character allows us to make a very different prediction that we normally would about another character, we find significant behavior correlation that might possibly give rise to a better classification basis. As I mentioned earlier, these preliminary results might allow me to form another classification basis which uses only the seemingly significant character states and limps the others together in a single void state which would not be used for evaluation at all.

good

GRAPH

Since this preliminary study did not include revisions before attempting the GRAPH computer run, it is not surprising that very little in the way of a classification system has been developed. The first pair that clustered did so at a similarity of .8200 and the last object to join when all other objects had clustered together was at .5400. In order to feel confident about a classification system, closely related objects should be .9000 similar or better and it should keep objects which are unrelated from joining together until .4500 or smaller. Since many of my characters distribute over one another with the same probability these results are to be expected. However, even at this stage one can find several significant groupings emerging.

why do you say that

Refer to page 8 and you will see the six groups which emerged by the time objects which were two-thirds similar or greater were included. Two of these groups are significant in terms of the present study. The first is the second group which contains every one of the objects labeled 200. These are the engineers. The group identifies itself when one examines several of the main objects in the group, for example 201 and 204.

201	204?
1. John J. Jones	1. John J. Jones
2. Engr.	2. Engr.
3. Block style Printing	3. Block style Printing
4. 120% Salary Increase Desired	4. 50-110% Salary Increase Desired
5. Quiet Constructive Activities	5. Quiet Constructive Activities
6. Lines Thru Mrs. Miss	6. Upstroke lines thru each
7. Married at age 24-27	7. Married at age 24-27
8. Words in the page margin 10-19%	8. Words in page margin = 10-19%
9. College Graduate or more	9. College graduate or more
10. 51-60 words on page 2 of form	10. 41-50 words on page 2
11. No tattoo	11. No tattoo
12. 13 or more negative words on form	12. 8-10 negative words
13. WHA rating of B	13. WHA rating of A
14. Oldest child in family with a sister	14. Youngest with sister
15. No scars	15. Scars

A quick check reveals the whole cluster as the conformist cluster and most of the responses fall into the categories which are most likely from a purely statistical point of view. The other interesting group is the 503, 504 group. The 500's are the Redfield Salesmen.

I am sure that in the light of some feedback about these individuals, more could be said about these groups, but for now their existence has been noted and they remain to be identified. Also it should be said that groupings with respect to several different parameters should be attempted individually and what I have here is most likely a grouping which contains these all at once. In order to have a greater resolution I will need to identify and seperate the actions of these several factors.

CONCLUSION

In a classification of objects one needs to know whether his characters have any relation to more abstract notions about them. Even more important, the cluster groupings should be checked against the criteria which the researcher uses to define them or against characters which are part of his intuitive evaluation of the objects. This kind of a crosscheck is feed back. In this study I had virtually no feed back and in this sense I was working in the blind. I learned much about the system and the problem from my investigations of the character states being compared, but I had nothing to really justify judgements about the cluster groups discovered by the Graph Program.

The results of this preliminary study indicate that much more could be accomplished in this area. The evidence which these results suggest indicates that with a more thoughrough study and with some positive feed back to cross-check and evaluate the results of each stage, a classification could indeed be made. Proper evaluation of the classification that results might suggest modifications of the input to include something other or besides merely the PH form.

Clearly the possibilities of a correctly working classification system can be made as great as one might desire. For instance if one were to carry this method to an advanced stage the following data could be collected: Information from employers after the applicant went to work for a period of time, an interview, the PH forms and perhaps a test. The information system thus created would provide a nearly exact picture of the individual. Through the use of the CHARANAL and GRAPH programs, a smaller subset of information bearing characters could be extracted from the PH form and an interview alone which could be used to make a valid prediction of the suitability of the individual.

January 1968

Gary D. Campbell

THE CORRELATION BETWEEN
INFORMATION THEORY AND EVOLUTION

prepared through
Independent Study
for

Dr. Olwen Williams
Prof. Biology, CU

INFORMATION THEORY AND EVOLUTION

I will present my discussion of the relation between Information Theory and Evolution in two parts. Part One will provide a very brief background of some of the concepts one needs before going on. Part Two will be the essential portion of this paper. In it I will attempt to provide a definition of life and possibly a hint as to a framework within which all of our scientific disciplines from physics to psychology and even the Social Sciences can be placed so as to interrelate. My approach will be qualitative and will raise more questions than it will give answers. My purpose will be to recognize and define the area of information theory as it applies to evolution. This will involve broadening the normal scope of information theory in an intuitive fashion to enable a complete relationship to be made.

PART ONE

ENTROPY

Entropy is physically measureable just like feet, seconds or degrees of temperature. The units that it is measured in are calories per degree Kelvin. This is how it is regarded by the physicist or chemist. Intuitively, entropy embodies the idea of randomness or the tendency for matter to lose its structure and pattern as its heat content increases. At zero degrees Kelvin the amount of entropy is defined to be zero, at any higher temperature it is the number of calories of heat which have been added to bring whatever it is we are measuring to that temperature, so that it is expressed as a quotient of the sum total of all the calories added divided by the temperature attained.

We note that ice at 273° K and water at 273° K differ in two respects: The heat content of an equal mass of water is greater than that of ice while the form and pattern present in ice has been lost to the much more random state of water.

The Second Law of Thermodynamics states: No change can spontaneously take place in a system which can neither give off nor receive energy from without, without increasing the entropy (or randomness) of the system.

TIME AND PARAMETERS

Time is a concept which we have devised to relate various things we can observe. It is a parameter to which we can refer many other things. One whould keep in mind that any of the things we speak about in terms of time contain this parameter either implicitly or explicitly and, until we have eliminated the parameter, the concepts we deal with are within our contrived framework rather than strictly within the framework of reality. As a parameter, time can be used to describe: Velocity, passage of events or their relative occurance, aging, and beginings and endings.

A parameter is useful when we want to speak of one thing apart from its relationship to the whole. We observe an object in motion and to describe it, we invent time. Now we can explain how its distance changes in relation to our parameter time relative to some frame of reference (which we also invent). Our frameworks of description are built through the use of parameters in order to contain a description of reality in simple terms. The fact I want to emphasize is that the parameters themselves are descriptive of relationship of two or more things that exist. The idea of relationship being of overall importance in later discussion, but existence and relationship are two distinct and different concepts. Relationships are implicit in reality but explicit in terms of parameters. Let us put things from our parametric frameworks back into reality, in this way they may be drawn out again into frameworks which make explicit other relationships that exist in reality, but which we aren't likely to see within the parameters through which we view reality.

For my purpose evolution and the growth or formation of an organism (pattern or intelligence) is also a time related process. Time is related to entropy. Once we have invented these two concepts or parameters, we find that with respect to the physical universe there are time related processes that can only occur in one direction. Since the universe is an isolated system, we can only expect the entropy of it to increase (Often stated: "The universe is running downhill".). This implies that all other time related processes must conform to this fact, and the parameter time and related processes must occur in a one-direction, irreversible, linear flow. This statement has often been called "Time's Arrow".

THE MAXWELL DEMON

In the latter part of the 19th century, James Clerk Maxwell proposed what has since been known as the Maxwell Demon. This hypothetical construction has been the motivation to relate the concept of entropy and energy to that of intelligence and information. It is a thought experiment which supposes two compartments, A and B, filled with gas. Between the sections is an ideal, infinitely thin, partition with a number of very small trap doors which can be opened and closed with no expenditure of energy. The demon who controls these doors is possessed with unlimited powers of vision and is instructed to open the doors whenever a molecule in A with more than velocity V approaches so that it passes through into B. Likewise whenever a molecule in B with less than velocity V approaches so that it passes through into A. Thus, given the situation where unlimited ability to perceive is coupled to the ideal situation of the two compartments, the temperature can be maintained at a difference on the two sides of the membrane and a heat machine could extract work from the system while no work has been done to the system.

What here seems like a paradox or violation of the laws of thermodynamics is actually resolved when one realizes the significance between work and intelligence gathering. Even with his unlimited powers of vision, the demon is faced with an original equilibrium in each compartment and he can not inform himself with the intelligence of particle motion from the available signal, even with unlimited powers of vision, since it is almost pure noise. The signal to noise ratio is the velocity and direction of any particular particle as one unit divided by the very large number of velocities and directions of all the similar particles in the foreground and background. To increase it, he will have to construct a new mechanism, because with even unlimited powers of vision no information can be had from a signal to noise ratio of nearly zero; he must provide his own information process and that will involve the expenditure of energy. In this way it can be shown (much more rigorously than I have here) that the collection of intelligence implies the giving up of energy. In information theory, an extension of the idea of entropy is made to include communication and intelligence gathering.

INFORMATION THEORY

In information theory several concepts are related in ways not in normal use. Information is a process. Literally it means to form into or to re-create out of something a given form or pattern. This form or pattern embodies intelligence or information content. When the information content is transferred, the actual process of transferral is called communication. Communication involves three steps: Encoding, Transmission, and Decoding. Two factors work to oppose each other in the communication process, and determine how effective the information process is, or what fraction of replication is obtained. These are noise and redundance. When the information process takes place, any deviation from ideal conditions is a factor which reduces the information content transferred and what is substituted in the place of intelligence which has been lost is noise. Noise can also add to the original intelligence and in this way also destroy parts of it. To counteract the effect of noise, the transmission redundancy may be increased and much or all of what is lost to noise may be regained. Leaving the actual mathematics out of it, another factor which is important is that of signal to noise ratio. If the signal to noise ratio were infinite, we could have an information process take place with the use of no redundancy in the communication, but as the ratio decreases to one, the redundancy must increase to infinity in order to get perfect replication. If the ratio were to fall below one, no information can take place at all, unless the receiver were to expend energy through some special process in order to change the signal to noise ratio.

PART TWO

LIFE AND LIVING THINGS

In speaking of life the first step to take is to make clear just what is meant by the term, this means defining boundaries. In a qualitative way one may set the boundaries for life to include all the things living or all of the matter currently making up living things and exclude all which is outside this realm. This means that a live animal is included, but a limb torn from a live animal (unless it is ingested by another animal) is removed from the boundary condition. At the same time any matter or energy which is a part of the metabolic process of a living thing is included in the term "Life" and any material, organic or otherwise, not actually included in a living metabolism is excluded. The term "Life" is the aggregate of all living things described in this fashion, but so far nothing has actually been defined. Unless we know what we are talking about to begin with, the idea of a boundary is still meaningless. It should also be emphasized that a boundary in this case doesn't imply isolation, there is of necessity much interaction across this boundary between Life and its environment.

Let us observe some of life's characteristics along another line. Life has the characteristic of decreasing its entropy and increasing its information content with the passage of time. Life absorbs energy and reabsorbs formations that it has already assembled, but which have been lost from it by removal from the boundary conditions.

Another characteristic of life is metabolism. Metabolism comes from a Greek word meaning change or exchange of material. This is the fundamental process which characterizes a living thing. It is through the metabolic process that a living thing can not only stave off increasing entropy, but actually bring about a decrease. With these concepts as a background I will now give a definition for life.

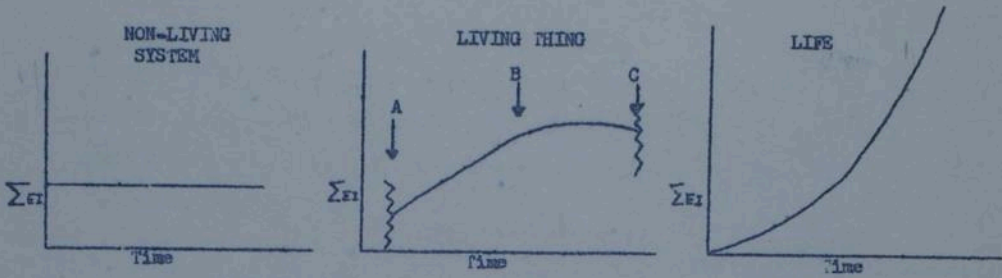
> A system may be called living, if upon the absorption of formation or energy (equivalent concepts) the entropy of the system decreases and the information content increases. Likewise any system with this property can be called living. The boundary condition which is implicit in this definition must include everything which is necessary to the system being described in order to bring about absorption of energy and low entropy formations, the expulsion of high entropy formations along with unnecessary energy, and the necessary information processes

involved. The unit or system to which this defintiion refers would be referred to in normal usage as a living thing or an organism.

With this definition, the former establishment of boundary conditions to describe the term life now has meaning also as a definition. None of this is meant to imply a contradiction to the Second Law of Thermodynamics. While Life is maintaining itself and even increasing, it is doing this only at the expense (in entropy) of its environment or often by direct intake of electromagnetic radiation. One might take all the energy absorbed by life, subtract from it the energy given off and the difference would represent mostly the order and formation contained by life or left behind as organic material.

What does seem to be a contradiction is that, speaking of the physical world apart from life, when energy is absorbed by a body or a system whose boundaries we recognize, by definition, its entropy increases. Because entropy means calories added divided by temperature attained. But entropy also means giving up information content or loss of form and pattern. What happens in the case of life? Let our system be a plant, it absorbs energy and grows. This is surely a contradiction! It is, in that we have changed the way in which we impose boundaries. If we use the same method of boundary imposition for the plant as we do intuitivly for the inanimate body, simply that of the volume of space which it occupies then in order to include all which is really relevent to the plant, we must also include its source of oxygen and minerals, carbon di-oxide and water. When energy is added we find that even though the plant itself has undergone growth and therefore a decrease in entropy, this has been off-set by the plant taking in material with low entropy and releasing some of it as waste with a higher entropy. This means that absorption of energy still causes entropy to increase except when we divide just the plant itself out of the environment and speak only about the plant. This is what we have done when we talk about a living thing or "Life".

Graphically we can better see how a non-living system, a living thing, and the total aggregate of what we call "Life" contrast to one another. In each case we let the total sum of the system's energy content and its information content equal $\sum EI$. In the case of the non-living system the proportions of each would fluxuate randomly, but the total would remain the same. In the case of a living thing the energy content would remain fairly constant but the information content would increase and in the case of Life itself both would increase. As a graph, each systems Σ EI plotted as a function of time would look like this:



For the case of a non-living system the sum remains the same and now it makes no difference how we define boundaries. For the case of a living thing (boundaries implicit) we have a curve somewhat more complex. Point A represents birth, a certain information content is already present. Interval AB represents the period of growth to maturity and reproduction; information content increases steadily. Interval BC represents old age, this interval would be very short for most animals and plants but in a few cases much longer than that of interval AB, this interval is marked by a leveling out of the information content increase and usually even a decrease will occur before death. Point C is death, the jagged lines at either end indicate entry and departure, into and from the boundary conditions.

For life in general the area under the curves for all individual living things summed together is relevent. The portion AB being the crucial part since during BC the organism will have off-spring alive and other factors enter in, in general the information content added by a living thing could be considered to be a straight line curve upwards with respect to time. This causes the graph for Life to increase more rapidly with time.

Besides the beauty of each graph being the mathematical integral of the preceding one, the first case is representative of the physical universe, the second of the information process and the third of evolution.

EVOLUTION

Now with a definition of a living thing and of Life we can see that the process of evolution is simply a kind of reaction going to completion. That it began spontaneously is the subject of two of the books in the reading list, but once having begun its continued process it is more a matter of observation than of theory. Just as entropy must increase with time, so must the process of evolution cause the subset of the universe called life to bring about order and decrease entropy with the passage of time. Much elaboration on this subject could be and has been done, but I wish merely to give some examples of how evolution has produced different types of information processes, each representing a new way of pattern formation based upon previously existing ones.

The first step in the evolutionary process, or where living became distinguishable from non-living, is believed to be that of the colloidal coacervate. Through an exchange with its surroundings (metabolism) it was able to maintain itself and reproduce a sufficiently precise replica of itself to begin the evolutionary process, the accumulation of more and more matter within the boundary condition.

Cells as we know them were eventually constructed and their takeover eliminated practically all trace of coacervate systems and other systems of borderline life, to become what is recognizeable as the fundamental unit of living material. Cells are characterised by an incredibly small scale complexity containing formations known as genes and chromosomes. These formations contain the intelligence which indicates the potential growth paths and characteristics of the cell and its progeny.

Multi-cellular organisms enable such specialized cells as nervous tissue to be formed. These cells (depending for their existence on interaction or metabolism within a larger organism with other cells) can collect and embody intelligence of a distinctly different level than that of the single cell. What is commonly referred to as information can now be collected from an organism's environment and stored for future use simply as "information".

Tools and artifacts are also part of the concept of life, as it is only when the organism producing them is taken together with these pattern formations that it can be studied in its entirety. The same is true with societies of organisms, no single organism within them contains all the complexity associated with its species, but rather the whole species (or sociological formations) must be studied or taken at once. All of this leads one to believe that information must have different orders of magnitude, or distinct levels upon which the information process can take place.

INFORMATION ORDER OR LEVEL

At some point it becomes necessary to distinguish between entropy and the lack of it and information and the presence of it. We can see that going from low entropy to high entropy; from intelligence to noise; order to chaos, that there exists a similarity. But the physical process of addition of heat and increase of entropy is reversible so that one may also take away heat and reduce entropy. When one has a particular pattern, however, and erases it, the constituents may be brought together again into that, or any of a number of different patterns, so that the original intelligence is lost and a new pattern created indicative of a new set of circumstances and causes which brought it into being. When arrangements of objects, whether they be pieces of chalk on a blackboard or atoms in a molecule are the basis of an information process, the concept of information is much the same. But we can see that some information processes depend upon previously existing ones to lend them order and give them their fundamental units and provide a medium in which to operate. For this reason I would like to propose a set of orders or levels on which information can take place and which each depends upon the establishment of the next lower one. Each higher order has also its whole spectrum of complexity but constitutes a kind of closed set. These might be labled as follows:

ORDER	DEFINITION
I	Sub-atomic particles, energy, fields and forces.
II	Atoms
III	Molecules and chemical reactions
IV	Single cells and biology
V	Multi-cellular organisms - structure and function
VI	Nervous tissue, behavior and psychology
VII	Societies and the social sciences - symbols, tools and artifacts.

ORDER		DEFINITION
VIII	The Abstract	

Each of those levels, as you can see, is a whole field of science unto itself and it is only my purpose to take note of some of these broad relationships, rather than to get myself irrevocably lost in trying to explore each one of them here.

INTELLIGENCE

Information as it is defined by information theory, involves the process of taking intelligence from one embodiment of it to another embodiment of it. This would be described in terms of communication as a three step process: Encoding, Transmission, and Decoding. When we speak of this or pattern or whatever it is that we lose when we increase entropy, we say that what we are talking about is intelligence. But what is intelligence? We could call it by a different name, information content, but the thing still eludes us. Whatever it is, it is clearly abstract, and whether or not a pattern, symbol, or anything in which the stuff of information can be embodied physically, the stuff must in somesense or another be.

Perhaps another aspect of this concept is the word definition. An arrangement can be defined or described without the possession of the set of elements to be arranged as long as the constitution of the set in mind will be conformable to the level or order of information. Embodiment, physically, of information content then is the realization of a potential. Calling something potential is simply to indicate a relationship between two things which are exchangeable, both of which are perfectly "real" but which have different qualities of being. In this way "Potential Information content" must also be a reality. We can now combine this with physical existence to define the total reality in which we actually exist. In other words reality is the union of the set of all that is physical with all that could potentially be arranged from that which is physical. Some of this latter set would have "meaning", it is not necessary for the arrangement to exist or the meaning to be known (which are actually the same things) but only to be recognized by us as a potential.

To begin at another point, we find information content embodied physically much as we can embody an obscure message in a metaphor. The two are in fact identical processes! The first using units of Level II for an information process on Level III, the second, using symbols, the units of Level VII for an information process on Level VIII. The important similarity is that it is the relation of the parts or elements we arrange one to another that distinguishes the intelligence they embody. These parts formed from what is available at one level - atoms - form together to embody all of the arrangements, processes and characteristics of chemistry to produce the single unit, a living cell, which is the basis of Life and succeeding orders which it enables.

INFORMATION DENSITY

In our corner of the universe the elements of Level I form almost exclusively atoms, sorfor our practical purposes the atom is the smallest building block. With atoms and forces we can have chemical reactions. Most of our pure science deals with what happens in this domain (Levels II and III and their interfaces). The frontiers of our science are pushing into the level on either side but after Level IV and to the greater extent on the fourth level, science is descriptive more than explanatory or predictive and we do not call science "pure" on levels greater than III. One thing which happens as we investigate higher levels is that the information density increases.

Information density in a qualitative sense, is the relative extent and order of the information process together with an idea of the intelligence involved per unit volume. It can readily be seen that after a certain magnitude of atoms are involved, complex chemical reactions can occur on a very small scale, but with a high degree of variation. How much volume does it take to represent a particular complexity? The less the volume, the greater the density and the higher the level of information.

For example, imagine the primeval ocean before the first bit of life had appeared, but when it had already accumulated a fair degree of simple "organic" materials. There would over the course of a long time exist quite a number of distinct chemical reactions involving these materials. But as compared to a single modern cell, the total complexity would scarcely be less and would probably be greater within the tiny volume of the cell.

The concept of finding the information density increasing over a period of time is identical with that of evolution. However, when seen this way, as an evolutionary process, we find that it takes more and more volume to form the sucessive building blocks - an atom, a cell, a human, society, and ???I In making a comparison of information density of two units on different levels, one must take the larger of the two and its volume and consider the smaller of the two redundant to the extent that it is necessary to obtain an equal volume. After the discovery of a new level of information by evolution it takes some time for it to arrive at the most efficient way in which it can embody the necessary complexity to form a succeeding level. We could take the size of an average atom, or cell, or ourselves as being representative of an approximate lower bound on the volume that it takes to embody the complexity of the respective levels. But the volume of society continues to grow and who knows what proportaons it will be necessary for evolution to take in order to embody the Abstract? Is the endpoint of evolution a complete embodiment of the Abstract by the physical universe? A sort of equilibrium between the tendency for the physical universe to increase entropy and the process of evolution to increase information content?

READING LIST

"An Essay in the Philosophy of Science" David Hawkins

A general work, this book is inspirational in places to almost anyone. Useful to my discussion were Chapters 8, 9, and 10, which dealt with much of my topic.

"What is Life?" Erwin Schrödinger

A particularly useful little book in tying the ideas of Biological Life and Physical Entropy together.

"Life: Its Nature, Origin, and Development" A. I. Oparin

One of the original works in the field of the origin of life, this book explains the concept of the colloidal coacervate and explains the processes by which life probably began.

"The Dawn of Life" J. H. Rush

Explains more for the layman many of the ideas included in Oparin's book and includes also other possibilities. Towards the end the author relates some of his more intuitive ideas about the present and future course of evolution.

"Scientific Uncertainty and Information" Brillouin

This is the latest work by one of the men instrumental in pioneering the field now called Information Theory. Its discussion contains less math than other discussions of information theory; it is useful for a more complete understanding of the field.

"Cybernetics and Biology" George

This book goes into man's attempts to understand life by trying to build models closer and closer to it. Some of the underlying concepts are explained in this relation including information theory.

POSSIBLE CLASSIFICATION OF CROSS-VALLEY MORAINES

J.T. Andrews

POSSIBLE CLASSIFICATION OF CROSS-VALLEY MORAINES

J.T. ANDREWS

Moraine is used in glacial geomorphology to describe a set of landforms that are associated with glacier margins. Commonly recognized types include; end, lateral and medial moraines. There is at least one other type that will be referred to as 'cross-valley moraines'. This moraine form is always associated with glacial deposition into a water-body (pro-glacial lake or sea). The moraines are closely spaced and are readily 'distinguished' from the other three major catagories. The c.v.m.'s were studied on Baffin Island, N.W.T. They can be divided into subsets using different criteria, such as shape, location of the moraine in the valley, relation of the moraine to former ice source, and then on various structural characters such as: strength of till fabric orientation and dip, mean grain size, pebble form and roundness.

The problem breaks down as follows; we can divide the moraines into subsets based on morphology and location, is this division meaningful in terms of differences in internal structure and composition? The matrix would comprise 80 individual sites times approximately 10 'characters'. The precise number of characters is not determined because I am not certain if the programs allow for missing data. Furthermore, it would be interesting to take each of the groups within 1, 2, and 3, and see how different a classification scheme might be insofar as the importance of 4 to 10 might vary.

SAMPLE NO. 80 x 11(?)

CHARACTER:

1) Morphology (1,2, ----))
2) Location (Valley)(1,2, ----)) -- Run each individually with 5 to 10 below?
3) Location (ice)(1,2, -))
4) Orientation strength(0 $\rightarrow \infty$)
5) Dip strength (0 $\rightarrow \infty$)
6) Dip type (1,2,3, ----)
7) Model dip up glacier/dip down glacier (ratio)
8) Stone form (1,2,3, ----)
9) Stone roundness, ratio round/angular
10) Mean size - Ø units

Class: Flowering Plants
Dr Rogers
Term project
Susan George

I am working on a classification of eight to ten species of the genus of lizards, Cnemidophora. Thus far I have examined and compared several individuals of each species and have chosen ~~a~~ ten characteristics as a means of separating out the species. These characteristics are,

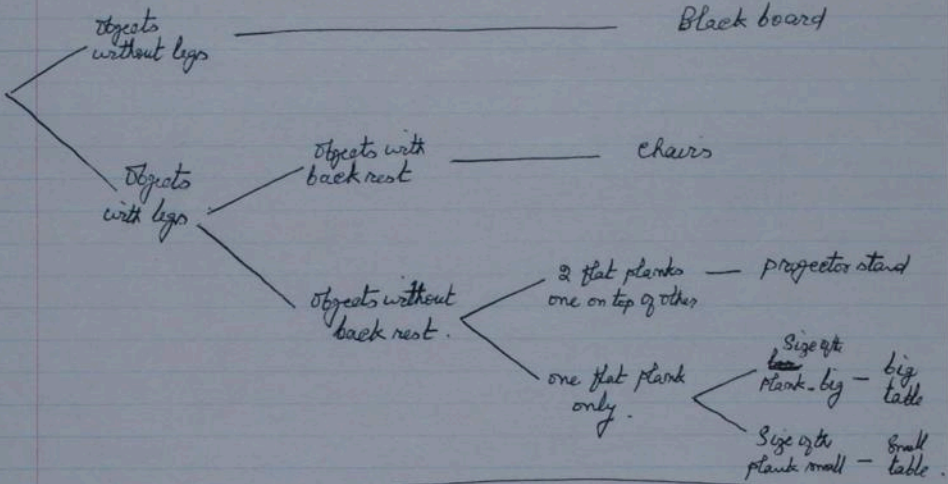
1. Number of femoral pores
2. Number of lateral superocular scales
3. ORG (granules down back)
4. Snout vent length - maximum
5. Temporal scale characteristics
6. Number and spacing of circum orbital semicircles
7. Supra anal scales
8. Gular fold scales
9. Post antebrachial scales
10. Color pattern in adults
    - a - Stripes - no spots
    - B - Stripes some spots
    - C. spots - some stripes
    - D. Spots - no stripes

Appen.

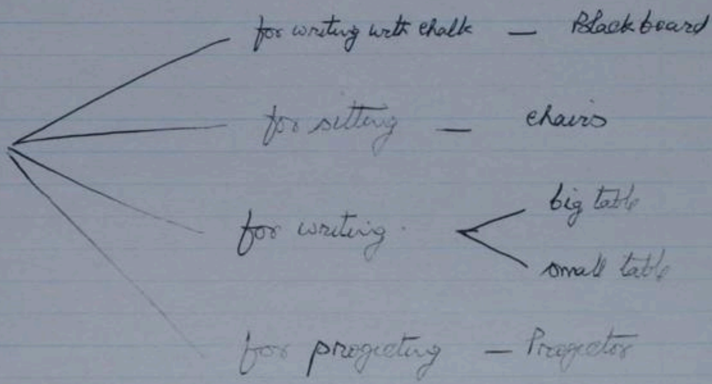
furniture includes – table, chair, blackboard, projector stand.

This classification is for the purpose of identification only

1) Classification on the basis of appearance.



2) classification on the basis of utility



Brij Kapoor

Objects made of wood, used for sitting and placing things on and having some legs — (4 here).

↓

Furniture

- Objects with flat surface on top — **Tables**
  - Length of top 15 ft. & Colour green — Large table
  - Length of top 6 ft. & Colour natural ↓ small table
  - Length 2 ft. & flat surfaces two & legs 8 Colour Grey
- Objects with an arm on the top & sitting space below — **Chairs**

Bob Egan

Genus: Furniturus

Classification of (furniture)

~~A~~ 1. Object made of wood - - - 2

~~B~~ 1. Object made of metal - - - 6

2. Color blond or ~~light~~ brown - 3.

2. Color greendark or gray - - - 5

3. Length 4 ft. or more - - - . 4
Flat top. 8 pieces →

3. Length < 4 ft., 25 pices - - - -
Furniturus chairis

4. ~~Flat top, 8 pieces of wood~~ - - -
Furniturus tabilis
var. blondus

5. ~~Table~~ objecT 10′ long, flat top,
darkgreen - - - -
Furniturus tabilis
var. verdis profundo

5. not as above, gray
16 pieces — Furniturus tabilis
var. grayii.

6. Color green ——
1½ ft. tall, round —— - - Furniturus
debrisis

6. Color gray metallic — round — ~~Furniturus~~ Furniturus

vedii

Bot. 431 V. Evenson

Furniture

I. Composed of wood ---- II

I' Composed of metal ---
a. Cylindrical, with sides about 18" tall - - - waste basket
a'. ~~Cylidr~~ Cylindrical, with sides about 5" tall - - - smoking receptacle

II Having four legs, ~~[illegible]~~ and a horizontal platform - - - - - - III

II' Without legs, hangs on wall, with a grey-green slate surface - - - Blackboard.

III The horizontal platform being low for the purpose of individual sitting, having an arm, color variable - - - - - Chair

III' The horizontal platform about 3' high, length variable, width variable, color variable - - - - - table

Aparna Mukherjee
5. 5. 68.

Study — Furniture

Class - Furniture

All objects in the room that helps furnishing the room should belong to the category of furnitures.

According to that definition there are several furnitures in this room like tables, chairs, black board, light holders, a ~~place~~ fire place etc. For the convinience of study I should only select those objects that are made up of wood only.

Class Furniture : Objects made up of wood.

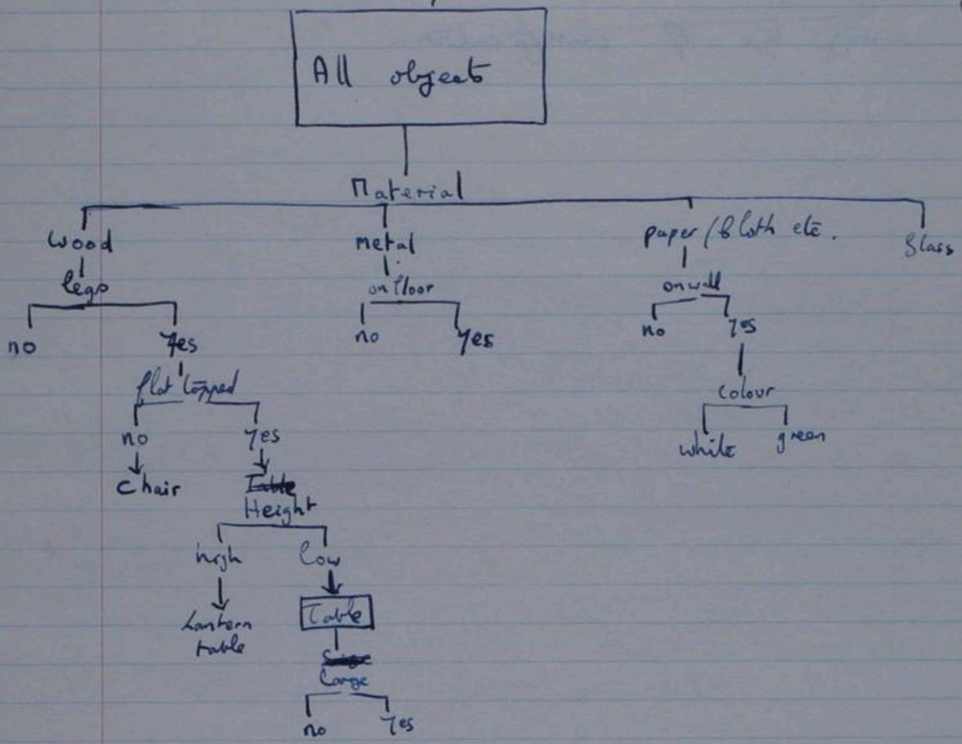
Group I : Chairs : ~~Pieces~~ Pieces of furniture used for sitting; piece of square or rectangular wood supported on four legs, with a high back; with one arm provided with a piece of wood; for writing, colors ~~natural~~. variable, brown or natural

Group II : Tables : Pieces of furniture, used for keeping things and writing; made up of piece of square or rectangular wood; supported on four legs; colors ~~can be~~ different, black or natural; sizes variable.

Group III : Desk : Piece of furniture, used for ~~reading~~ holding books' etc., Color natural

An hierarchial classification is useful but is not the only type of classification.

Classification of Furniture



Class - Furniture   Susan George

1.a ~~A~~. Objects composed of wood - mainly - - - - - 2
B ~~B~~. Objects composed of metal - - - - - - - - - 4

2a Objects used for sitting and writting - - - - - - desks
B Objects used only for writing only - - - - - - 3

3a Objects ~~for use~~ - natural wood color - - - Smalltables
B Objects - green - - - - - - - - Largetable

4a Objects on the floor - - - - - - - - - - 5
B Objects on the walls or ceiling - - - - - 7

5a Objects for descarded material - - - - - - 6
B. Object for ~~cleaning~~ washing - - - - - - - Sink

6.a Objects for cigarettes only - - - - - ash trays
B Objectes for trash - - - - - - - wastebaskets

7.a Objects for light - - - - - - - - - ~~light~~ 8
B Objects for showing movies - - - - - screen

8a ~~Objects for~~ light adjusting object - - Window blinds
B luminiscent objects - - - - - lights

G D Campbell

Furniture in Room
HOROZONTAL SURFACE PLANE — CHARACTER
NUMBER OF PLANES
AREA OF PLANES
WHETHER FLAT OR WAVY OR TILTED
COLOR OF MATERIAL INCLUDING THE PLANE

A TAXIMETRIC STUDY OF SOME SPECIMENS OF *ASTRAGALUS*

Botany 431

Virginia Evenson

Virginia –

I'm sure that you could have done a better job had you had more time. You seem to have the general ideas in hand, but didn't quite express them accurately. I hope you can continue to investigate the technique for application to your own interests.

D.J.R.

Course B
Talk B-
Paper C

The problem of how to classify things is one which has intrigued and puzzled workers for many years. The methods which have been devised and employed in solving this problem are almost as numerous as the numbers of objects to be classified. Within the last few years a new field has arisen, one which employs a new method to help solve the problems of taxonomists. This method uses mathematical analysis and computers to help ~~give~~ analyse information, show relationships, and aid in the classification of objects. This new field is taximetrics, the methods of which were used in this study.

A study was made of some dried specimens of the genus *Astragalus*. All of the specimens were collected in and around the Fort Collins, Colorado area. It was suspected in the beginning of the study that some of the specimens were not of the genus *Astragalus*, but were a member of a closely related genus, *Oxytropis*. It was decided to include these specimens in the study also to see what relationships, if any, could be determined between the two genera.

The genus *Astragalus* is a member of the Leguminosae (pea family), which is divided into three subfamilies, Lotoideae, Mimosoideae, and Caesalpinioideae. The subfamily Lotoideae, to which *Astragalus* belongs, contains the majority of legumes of temperate regions of both the northern and southern hemispheres and includes such

Should you include a reference for these descriptions? →

genera as Trifolium, Lupinus, Oxytropis, Pisum, and Medicago.

Astragalus is a perennial (sometimes annual) herb, with roots being rhizomes or taproots. The leaves often consist of leaflets and are variously pubescent. The flowers vary in color from white to cream to purplish, have diadelphous stamens (nine united and one free), and a glabrous style with a terminal stigma. The fruits of Astragalus are pods which are diverse, being sessile to having a stalklike base, and varying in texture from membranous to woody.

The method used in classifying was that of converting descriptors into measures of taxonomic similarity, and on the basis of the similarities, dividing the objects into clusters.

The specimens in the study were described by choosing characters which were thought to be significant and coding each object into its corresponding state of the character. The choice of characters is an area in which the worker can exercise his judgment and skill as a taxonomist. Meaningful characters must be chosen so as to show relationships between objects, and a thorough knowledge of the subject matter is prerequisite to making these selections.

The characters which I chose to describe my specimens were as follows:

I would have thought that considerably more and varied characters would have been apparent to you!

K_1 Flower color

1. cream
2. purplish
3. cream with purple dot

K_2 Length of pod

1. 5-10 mm
2. 10-15 mm
3. 15-20 mm
4. 20-25 mm

K_3 Texture of pod

1. papery
2. leathery
3. woody

K_4 Leaflet

1. sessile
2. not sessile

K_5 Pubescence of leaflet

1. leaflets densely pubescent
2. leaflets sparsely pubescent

K_6 Length of stem

1. 10-20 cm
2. 20-30 cm
3. 30-40 cm
4. over 40 cm

When all objects were coded into the various states, the coding form was analyzed by a computer. This analysis tells the worker the amount of information contained within the character, the probability of the states of the characters, and the conditional probability distribu-

A more comprehensive discussion would be appropriate here.

objects, Not descriptors are compared here!

tions of the characters when compared with the remaining characters. This information is numerically presented and is most useful in telling the worker the correlation and interdependence between descriptors. If one character shares little or no information with the other characters, it may add much to the classification of the objects. On the other hand, if two characters share the same information, this would indicate to the worker that one of the characters is not necessary and could be revised or eliminated from the study. It is up to the worker to interpret the numerical results of this analysis and determine if the characters which were used in the study are meaningful or if they should be altered or perhaps eliminated from the study.

Pairs of objects were assigned a similarity measure based upon the comparison of the descriptors. This similarity measure is expressed numerically, with objects having a high similarity measure being more alike than objects with a low similarity measure. This similarity measure was then used to convert the objects into clusters for classification. Each cluster was given a numerical measure to express its isolation from all other clusters. The larger the numerical value (C-value), the smaller and more numerous become the clusters, so that objects within each cluster are very similar to each other. As the numerical value of the cluster becomes smaller and smaller, the classes become larger and fewer. The clusters are graphically shown (Skiline) in reference to the objects and the corresponding C-values for the clusters. At each C-level, cluster memberships are shown. As the C-value drops, more and more objects are connected and added to the membership of the cluster. The objects which are joined in at the lower values are more loosely connected in the membership, that is, they show less similarity than those added at higher values. The difference between two successive C-values is called the moat of C, which can be thought of as the stability of the cluster.

By this method, the phenotypic similarity between objects is shown, the phenotypically similar objects being put together very early (at a high C-value) and the phenotypically dissimilar objects are separated.

RESULTS OF CHARACTER ANALYSIS

Each character state was compared with the other states to show the probability of the one state when the other state was known. For example, when the specimen had woody pods it was found to always have cream colored flowers with a purple dot, when the flower was cream colored, the pod was always papery in texture, and when the plant had purplish flowers the pod was always leathery. The correlation of these characters in this way might

Redundant information

Provides taxonomic

structure!

why?

suggest that one or the other of the characters could be changed, but the fact that information was lacking in some of the specimens indicated to me that there should be more specimens from which to draw a more valid conclusion. I might add at this point that when information was missing all of the objects were not included in the comparison . This fact must be taken into consideration when analyzing the results.

A better conclusion can be derived from the comparison of flower color and the attachment of the leaflet. Here it was found that cream colored flowers were always correlated with leaflets with petioles and the purplish flowers were always associated with sessile leaflets. Another problem that could arise here and should not be disregarded is the fact that color in dried specimens often fades and what was purple in the living specimen could fade to white in the dried specimen. For this reason, it is good to take this kind of data from the plant when collected so as to not make errors. This fact should be taken into consideration before making absolute statements about correlations.

The pubescence of the leaflets in plants with papery pods was always found to be dense, while those plants with leathery pods always possessed leaflets which were sparsely pubescent. It was also found that leaflets with dense pubescence were always sessile, and those

It might be appropriate to cite an authority for this formal conclusion

 Singular

which were sparsely pubescent were sessile 83% of the time. Since all objects were included in this comparison, the information given by these figures is significant and perhaps indicates that character four does not contain much information and probably should be altered or eliminated.

RESULTS FROM SKILINE

That there were indeed two taxa in this study can be seen from the Skiline graph. There are two distinct clusters which do not become joined until a C-value of .4800. There were only two connections between the different taxa at this level. The taxa are well defined as separate groups at a high level with a C-value of .8000.

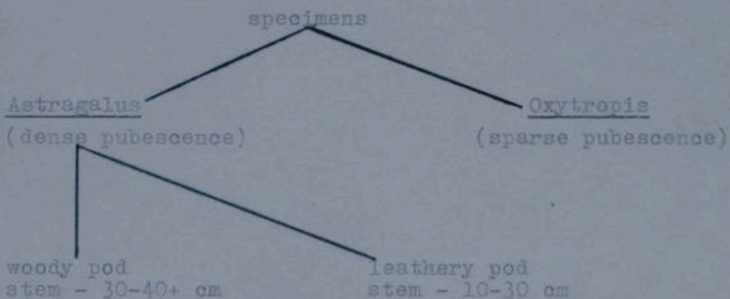
From the comparison of the coding with the two different groups, the Oxytropis specimens are separated from the Astragalus on the basis of the pubescence of the leaflet. All the Astagalus specimens except one are coded as having densely pubescent leaflets. The one Astragalus specimen that was coded as having densely pubescent leaflets is the one specimen which made connection with the Oxytropis specimens at the C-value of .48000. When this one specimen was re-examined, the leaflets were found to have dense pubescence rather than sparse pubescence, so the separation of the two taxa is valid and would have been at an even lower C-value had not could you have suspected this from the Character Analyses?

this error been made in coding.

The Astragalus specimens can be further divided into two groups based on the texture of the pods and the length of the stem. One group has leathery pods and stems which are 10-30 cm in length. The other group has woody pods and stems which are 3-40+ cm in length. These two groups cannot be further separated on the basis of the characters used in this study.

Ant.?

On the basis of the results I have constructed the following key for the specimens of my study.



This study has by no means been an exhaustive examination of the genus Astragalus in Colorado or even in a particular area, but has given some insight into how the taximetric method works and can be applied to the field of taxonomy and the problems of classification.