Limitations of Archaeological Inference: An Information-Theoretic Approach with Applications in Methodology
Author(s): John S. Justeson
Source: *American Antiquity*, Vol. 38, No. 2 (Apr., 1973), pp. 131-149
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/279360

LIMITATIONS OF ARCHAEOLOGICAL INFERENCE: AN INFORMATION-THEORETIC APPROACH WITH APPLICATIONS IN METHODOLOGY

JOHN S. JUSTESON

ABSTRACT

A framework is established for the application of information-theoretic concepts to the study of archaeological inference, ultimately to provide an estimate of the degree to which archaeologists, or anthropologists in general, can provide legitimate answers to the questions they investigate. Particular information-theoretic measures are applied to the design elements on the ceramics of a southwestern pueblo to show the methodological utility of information theory in helping to reach closer to that limit.

Department of Anthropology
University of California, Berkeley
September, 1971

IN RECENT TIMES there has developed within archaeology what has been called a *revolution*—a reorientation of the goals, an explicit definition of the assumptions, and an overhaul of the methodology. The stated attempt is to make archaeology a predictive behavioral science dealing primarily with long-term cultural change. Another and older perspective treats archaeology as a method of reconstructing the social and cultural histories of specific human populations. The 2 really seem only to be poles on a continuum of research commitments held by American archaeologists, few of whom seem to truly reject the latter and perhaps even fewer to reject the former description as characterizing the proper interests of their colleagues. Both perspectives can be expected to survive through the next several years at least; both share the same means of data retrieval; both have the capacity to treat the same data. A good summary of the current situation has been presented recently by Thomas King (1971).

This article does not really fall under the rubric of either "new" or "traditional" archaeology. Rather, it is an attempt to ascertain the inherent limitations on the ability of either perspective to attain its goals; to present ways to determine in any given archaeological situation the further limits imposed by the particular circumstances of the formation and excavation of the archaeological record; and to suggest particular tests which may be applied in helping the archaeologist to reach those limits. *Inherent limitations* is intended to mean those limitations which arise from the nature of the archaeological record itself and from the manner of its formation. These limitations can be dichotomized into 2 basic categories: (1) limitations imposed by the degree of preservation of culturally significant remains and by the skewing of their relationships through time until their recovery; and (2) limitations on the interpretability of archaeological data for the cultural descriptions. The first presents problems because of the possibility of loss of crucial information, or of its redistribution into other than its original contexts; the second is a problem because of the incomplete nature of the identification of the material consequences of cultural behavior with the cultural forms that produce them. It is the first of these areas of limitation that this paper attempts to deal with. Study of the second is now also under way.

It was stated above that traditional archaeology is concerned with the reconstruction of ancient cultural systems while the "new archaeologist" looks for explanations of cultural change that can be used to predict particular changes in particular situations. The basis for change sufficient to shape new cultural processes must underlie any valid explanation of a cultural process, and therefore evidence of the motivations of the groups of people involved is required. However, in the absence of consciously transmitted and decipherable symbolic statements of motivation in the particular instances of interest to the archaeologist, there can be no direct material evidence of cultural change beyond associations of temporal priority, contemporaneity, or temporal overlap between various cultural systems which leave material remains. Therefore, only a succession of cultural descriptions can be transmitted through the archaeological record; further interpretation, whether predictive or retrospective, must be of a higher order of abstraction than mere extrapolation from known data. Here, then, is another area of overlap between traditional and new archaeology—both must reconstruct cultural systems from their primary data, and upon this they may base their further investigations or their subsequent tests of prior hypotheses.

Granted these opening statements, we may put the immediate goals of data-gathering as the construction of aspects of the cultural system, whether they be external factors affecting its operation such as ecological interactions or internal factors such as economically-based social differentiation; the ultimate research goals of the investigation may, of course, be quite different. It is at this basic level that we shall attempt to present ways for estimating the limits of the effectiveness attainable in the archaeologist's efforts.

The approach to be taken to this problem is information-theoretic.

> Information theory is one of the youngest branches of applied probability theory. It started with what is now the classical work of Claude Shannon in 1947-48 and has since grown considerably. A new impetus to the subject was given by A. N. Kolmogorov in the year 1958 when he discovered the close relation between the notion of entropy and the classification problem of classical systems. The subject has applications in many other branches of science as well [Kotz 1966:iii].

To archaeology it is a very new field. Although the entropy function, or measure of information content, was used by Arthur Saxe as a measure of diversity in social organization (Michael Schiffer, personal communication), this use is not strictly information-theoretic any more than is the use of the same function as a measure of diversity in ecological studies (Margalef 1968) or of disorder in thermodynamics (Sonntag and Van Wylen 1966). Information theory proper has apparently been previously applied only by Frederick Gorman, who utilized tests derived by use of information-theoretic notions in an attempt to construct a consistent taxonomy of projectile point types and to delimit the range of archaeological contexts in which it was applicable (Gorman 1970). Such uses of the concepts of information theory in methodology seem important in that they allow an increase in precision in the construction of categories that form the basis of subsequent interpretation of the archaeological data. They are not the only concern of the present article, however. It is notions of the structuring of the transmission of information that form the basis of this study; their application to archaeological modes of interpretation, it will be shown, can be expected to provide some measure of the efficiency of archaeological data and of the modes of interpreting them for the solutions of problems of interest to archaeologists. The first part of the paper will therefore be dealing more with properties of channels for the transmission of signals and their connections with the source and with the receiver, and with the linking of these concepts to archaeological situations. It is a frankly speculative attempt to explore the utility of information theory to the study of the potentialities of archaeological research. [The mathematical concepts which form the basis for the discussion of the basic concepts of information theory as such and of the section on channel classification and channel properties are from Kotz (1966) and Wolfowitz (1961). The measures used in the sections on system dynamics and binary coding applications were from Herdan (1966).]

PART I—A THEORETICAL FRAMEWORK

Basic Concepts of Information Theory and Their Archaeological Correlates

Information is measured formally on the basis of attribute contrast. Where there is no contrast there is no distinction, and a distinction between things is at the base of our everyday idea of what information is—the organization of distinctive qualities into integrated wholes. The specific measure of information used in information theory is the *entropy function*, named after a similar formula from statistical thermodynamics. It is given by the expression

$$H = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

where H is the information content, n is the number of distinctive attributes in the system, x_i is the i^{th} attribute, $p(x_i)$ is the probability of occurrence of the i^{th} attribute, and $\log_2$ is the logarithm function to the base 2. Log_2 (for any quantity of y) can be calculated from the values of $\log_{10}$ y, which can be found in tables of the common logarithm, by the formula $\log_2 y = 3.321928 \log_{10} y$. As an example, if the system under consideration were an alphabet, the letters would be the attributes, n would be the number of letters in the alphabet, and x_i would be the i^{th} letter of the alphabet. Clearly, the order given to the attributes by their indexing is of no consequence in determining the amount of information in the system; it is simply a helpful bookkeeping device to keep notation from becoming too cumbersome.

In practice, of course, information must be transmitted to be of significance, and information theory does deal with the problems of transmission. The general scheme may be explained by reference to a telephone conversation. There is a *message* in the mind of the speaker to be transmitted to a listener. It is not the message itself that is transmitted, however; the speaker first puts his message into symbolic form by means of language. The spoken sequence is a *signal* that is *input* into the *transmission channel*, which in this case is the whole electronic complex that forms the telephone system. From this transmission channel another signal is *output*, in this case approximately the same as that which was input. The output signal is then reformulated into a semantic unit which is the *output message* of the process. This basic formulation is summarized in Fig. 1.

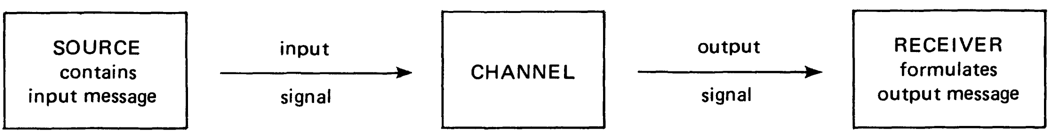

Fig. 1. Schematic representation of information transmission.

Inherent in this description has been the notion of a *code* by which semantic components are linked with symbols used to transmit them. In language, the code may be thought of as the transformational grammar which connects deep structure with surface structure, at least from one theoretical linguistic perspective (Postal 1971). In the given example, *encoding* also occurs at the link between the input signal and the channel and *decoding* at the link between the channel and the output signal in the transformation of audial signals into electronic impulses and then back again into audial signals. It is this second usage that is important in information theory normally but not to the problem we hope to explore.

Another factor not mentioned in the above description but which is familiar to anyone who has ever used the telephone is the presence of *noise* in the channel, an interference not directly related to the message whose signal is being transmitted, but which does itself form a part of the signal. This noise can be random, due to random fluctuations in the electronic impulses, or it may be systematic, for example, as the result of interference from other signals being transmitted. Basically, noise is any addition to the signal that is unrelated to the message. *Distortion* is also familiar in the variation in volume, pitch, or tone of the signal, and can be introduced either in the channel during transmission or at the source, that is, at the point of emission of the input signal to the channel from the speaker. It is basically an alteration in the relationships of the elements forming the transmitted signal.

With these basic concepts in hand, we may go forward to their formulation in terms that are meaningful to the archaeological situation. *Information* in this case will be a particular type of information, namely, the description of the cultural system. The contrasts which form the basis of our information on cultural patterns are based on the material artifacts which are recovered archaeologically and their patterned associations.

The *input message* transmitted archaeologically is the synchronic description of an aspect of the cultural system whose functioning was responsible for the material remains that the archaeologist studies.

The *input signals* are the material consequents of cultural behavior and of factors affecting that behavior.

The *archaeological channel* is perhaps a greater abstraction than the other concepts in application; looking back to the telephone example should help to clarify it somewhat. In that instance the channel was the electronic system that accepted the input signal and delivered the output signal. In our case, it is generally the earth itself to which the input signals are committed, and it is from the earth that they are drawn. There are exceptions to this generalization; if astronomical events are important to a study, for example, as they were in a recent study of the orientation of Complex A at La Venta in Mexico (Hatch 1971), then not all of the data were drawn from the earth. Such exceptions, however, are probably insignificant quantitatively, the great majority of data in archaeological studies having been gathered from the earth. The question is of no great interest except as clarification; in the exceptional cases we simply have more than 1 channel operative, and these can be treated separately and linked to get the results being sought. There is no need for insisting upon a single channel for the reception of any and all messages of interest to the archaeologist. However, the choice of channel can be important with respect to the results expectable from that channel, since channel properties have strong effects upon signal transmission and reception.

The *output signal* is the material assemblage that the archaeologist digs up and the associations that he notes; these include the surviving materials or the traces of them which formed the original output signal, and the other materials subsequently introduced. These subsequent materials, together with subsequent cultural or natural disturbances caused by later human or non-human visitors or events, are at least some of the factors that introduce *noise* into the archaeological channel. This includes the absence of once-present artifacts due, for example, to natural deterioration, to potting, or even to the manner of excavation. *Distortion* results from cultural "misbehavior," that is, from the deposition of material artifacts resulting from unpatterned deviations from the standard, causing the transmission of an inappropriate signal; and from post-depositional activity disturbing the associations or even the existence of attributes.

The code joining signal to channel is a trivial construct: the input signal is the material assemblage laid down, and their laying down is the "code" that commits them to the channel. The "code" by which the output signals are recovered is, likewise, the process of recovery.

The code joining the human behavior to its material consequents as committed to the earth is a code which is not at all trivial; it is the crucial concern of the archaeologist, since it is through it that cultural description is attempted.

The above description of the total communication process becomes somewhat more complex if we reconsider the input signals as being primarily the inputs and outputs of a communication process that was operative within the original culture. In that case it becomes clear that the archaeologist is in the position of the code-breaker tapping a channel with whose code he is not fully familiar by means of another channel. It is an elaboration which will not affect the rest of the discussion, but which should be borne in mind.

Channel Classification and Channel Properties

There are several aspects of information theory which, given the equivalences outlined above, seem to have potential for the exploration of the limitations on archaeological inference at the "nature-of-the-data" level.

Since channel properties are basic to the results retrievable from any transmission, it is important to know what kind of channel we are dealing with, because channel properties vary across channel types. Channels may be classified in any way, the system of classification being a priori arbitrary; the utility of the classification depends upon the particular application to which it is put. However, certain types of channels have been distinguished because of their simplicity and because they are of wide applicability. The basic distinction is between memoryless channels and channels with memory. In a memoryless channel, any transmission is unaffected by any other transmission, and the elements of the input signal are unaffected by the other elements of the signal; in a channel with memory, past history does have a bearing upon later transmissions. Another distinction is between finite and infinite input and/or output alphabets. The input alphabet is the set of symbols drawn upon to form the input signal, which the output alphabet is the set of symbols drawn upon for the output signal. The symbols of these alphabets have been mentioned earlier in this report as "elements."

For the archaeological channel, since the number of meaningful attributes in any material system is finite, and since both the input and output signals are material systems, then we can say that both the input and output alphabets are finite. A channel with this property is called *discrete*. In a discrete memoryless channel, the alphabetic symbols must be transmitted and received independently of each other. Our channel, then, must be discrete and with at least finite memory, for, while memoryless channels do exist archaeologically, their signals are undecipherable from our vantage point since, being independently transmitted, they show no differentiation or special clustering on any level; independence implies only random associations. In general, one would therefore not expect such signals as would be transmittable over a discrete memoryless channel to reflect messages of importance within the original cultural system since they can show no differentiation, hence no special differentiable function, hence could perform no differentiable task within the total cultural system. As such, they may be classified either as representative of individual idiosyncratic behavior or as associated with ideas or behavior having no further remaining material consequents.

Any signal transmittable over a discrete memoryless channel is transmittable over a discrete finite-memory channel if the symbols transmitted are different from those transmitted dependently over the channel. If they are not different, then we may make a distinction between the knowable and the unknowable and only those channels which have at least finite memory would then need to be considered. The channel must also have at most finite memory, however. A Berkeley mathematician, for example, finds the truth of the theory of the evolution of life from non-living matter to be historically unknowable because of the accumulation of noise through time (J. Rhodes, personal communication). Because of this accumulation, we can put an upper bound, however large, on the length of time before all memory of a transmission is blotted out. The discrete finite-memory channel is thus an appropriate characterization of the archaeological channel. As a qualifier, however, it should be noted that in the discrete finite-memory channel, memory is assumed to be unimpaired through some finite number m of transmissions and to end abruptly thereafter. In fact, the channel with decreasing memory of past history would probably be a better characterization. However, requirements of independence of signal probabilities over time—this property is called *stationarity*—form the basis of the theory of decreasing-memory channels (Wolfowitz 1961; Kotz 1966; David Blackwell, personal communication). While these requirements are not necessary to the theory, the investigation of these channels would be more difficult without it. Human behavior, however, does not satisfy these requirements (Quastler 1956).

Knowing what type of channel is being dealt with is simply a preliminary to the study of channel properties. It is important to know now what properties of channels are important to consider archaeologically. The major thing to look at here is the code. The code is defined mathematically as a system of N ordered pairs consisting each of an input sequence u_i of n alphabetic symbols and a set A_i of output sequences, where the N sets of output sequences do not have any members in common, and where the probability that the sequence received when u_i is transmitted will be among the members of the set A_i will always be greater than or equal to $1-\lambda$, where λ is greater than 0 and less than or equal to 1. In symbols, it is "a system

$$\left\{(u_1, A_1), \ldots, (u_N, A_N)\right\}$$

where the u_i are *n*-sequences, the A_i are disjoint sets of *n*-sequences and

$$P\left\{v(u_i)\epsilon A_i\right\} \geqslant 1 - \lambda, i = 1, \ldots, N$$

. . . we shall call it a code (n, N, λ)" (Wolfowitz 1961:51-52). The expression $v(u_i)$ represents the signal received when u_i is sent, while the term *n-sequence* is an input or output signal of length n.

The parameters n, N, and λ therefore specify the code for the channel. Without such a code we cannot really speak of information being transmitted or received, hence there is really no basis for speaking of the existence of a channel. We can find out if there is a code for the archaeological channel by finding if values we compute for N, n, and λ are consistent with the requirements of a code for a discrete finite-memory channel. In particular, the value of N is related to that of n by the formula $N = 2^{n(C-\epsilon)}$, where C is the *channel capacity*—a measure of the ability of the channel to transmit information—and ϵ is a positive constant. C may be determined, often only with great labor, from the relation

$$C = \max_{\pi} \left\{ \sum_{j} \left[\sum_{i} \pi_i w(j \mid i) \log_2 \sum_{i} \pi_i w(j \mid i) - \sum_{i} \pi_i w(j \mid i) \log_2 w(j \mid i) \right] \right\}$$

where $\pi = (\pi_1, \ldots, \pi_k)$ is any probability distribution, $w(j \mid i)$ is the probability of receiving j if i is sent—j can be null—and k is the number of elements in the input alphabet. A simple method for calculating an upper bound for the capacity may be found in Helgert (1967); it would require too much discussion for presentation here.

If the empirically measured parameters are not consistent with the relationship between them that is required by the theory for a given material or behavioral system, then the data by which that system is to be interpreted cannot have a consistent susceptibility to decoding; that is, there will be no basis for deriving a coherent archaeological interpretation of the data that will accurately reflect the prehistoric situation. Thus, the question of the existence of a code is one of primary importance for our considerations.

PART II—APPLICATION OF INFORMATION-THEORETIC MEASURES

The present section marks an important departure from the earlier discussion. While Part I was concerned mainly with channel properties and coding requirements for application to the study of archaeology and of archaeological inference, and was therefore mainly theoretical in nature, the present section will deal with the application of specific information-theoretic measures to a particular archaeological situation. At the same time, it will feed back into the final discussions of the last section. The applications will not be to the interception of information on some aspect of human behavior, such as an economic system, since the material results of such systems are manifold, would require much more intensive investigation than time allows at present, and would be better treated initially by an ethnographic study. Rather, these applications will treat the interception of formal information—information measured on the basis of formal contrasts by the entropy function—in an internally coherent material system.

The particular material system chosen for analysis was the system of ceramic design elements; we refer to them as constituting a system because they are presumed to reflect an aesthetic, ideological, representational, or functional system or an interlocking set of them, so that at least some of the relationships embodied in those systems should be reflected in relationships among the design elements in some way. This system was chosen because of the ready accessibility of an unpublished situational tabulation of design elements by Frederick Gorman of the Southwest Archaeological Expedition. The design elements are all those occurring on the ceramics recovered during the 1970 excavations of the Joint site, a 36-room pueblo in the Hay Hollow Valley of east central Arizona, provisionally dated to between A.D. 1100 and A.D. 1300. The initial step in setting up the entropy analyses was the extrapolation of the number of design elements actually used at the Joint site from the number found in excavation. Distributional patterns were not considered in these applications, either as attributes along with the design elements or as conditional factors affecting the frequencies of occurrence of design elements. The reasons for this exclusion seemed sufficient to justify it: (1) distributional patterning is a different order of phenomenon than the attributes which it patterns, so the method of alphabetic extrapolation, to be described below, will not give good results if they are included; (2) distributional patterning of other attributes is generally on larger groups of attributes, so that it implicitly results in the creation of a qualitatively larger system than simply the system of attributes of material artifacts that we are to consider as an example; (3) patterns of association can be ignored altogether in entropy computation, just as context of occurrence is ordinarily ignored for the purposes of computing information content of linguistic alphabets. The value resulting after this exclusion gives the entropy independent of context. Taking the distribution into account should result in somewhat better values if the totality of the relations of the given system with others outside it are important. The difference may not always be great; for example, entropy in printed English based upon word-length frequencies gives $H = 2.274$ uncorrected, while contextual considerations result in a value of 2.3 (Herdan 1966:303). While distributional patterns are therefore not necessary to the present study, a contextual conditioning for material systems is the only way of approaching the problem of behavioral systems archaeologically. Where behavior is the focus, then, conditioning on context would be important.

Extrapolation of the Prehistoric Distribution of Design Elements

The extrapolation technique to be described is a general one that can be applied to find the expected number of attributes that functioned originally in any well-defined attribute set. (By *well-defined* is meant an attribute set in which all attributes are comparable. For example, punctuation and letters are not comparable attributes in printed English, nor are main signs and affixes in Maya hieroglyphic writing.) It is based on a technique that was used to extrapolate the number of alphabetic characters in the type-fount of the Phaistos Disk (Mackay 1965). In the article cited, the technique was based upon small samples and involved a linear relationship between the variables involved in the extrapolation. Here, however, we make no assumptions about sample size and do not impose in advance any particular relationship between the variables; larger samples may, in general, be expected to result in a curvilinear relationship, however.

The first step in using this technique is to tabulate the values of N_t, the number of attributes occurring exactly t times in the sample, where the values of t range from 1 to T, and T is the maximum number of times any attribute occurs. Then M_t, the number of attributes occurring at least t times, is calculated and tabulated. M_t can be calculated from the formula

$$M_t = \sum_{i=t}^{T} N_i$$

M_1 is the number of different design elements occurring in the sample.

Finally, L_t is calculated and tabulated. Its formula is

$$L_t = \sum_{i=t}^{T} M_i$$

L_t has no meaning except when $t = 1$, at which point it is the number of occurrences of attributes in the sample.

These values can then be plotted on semi-logarithmic (log-normal) graph paper in such a manner that the relationships plotted are between $\log M_t$ and t, and between $\log L_t$ and t. From the curve obtained, a value at $t = 0$ can be extrapolated to give M_0, the number of attributes occurring at least 0 times in the sample—the total number of attributes in the system. Since $L_0 = M_0 + L_1$, the extrapolation of L_0 provides an alternate check on M_0.

This procedure was carried out for the design element data from the Joint site. The sample size L_1 was 2317, with $M_1 = 100$ being the number of different design elements recovered. A generally

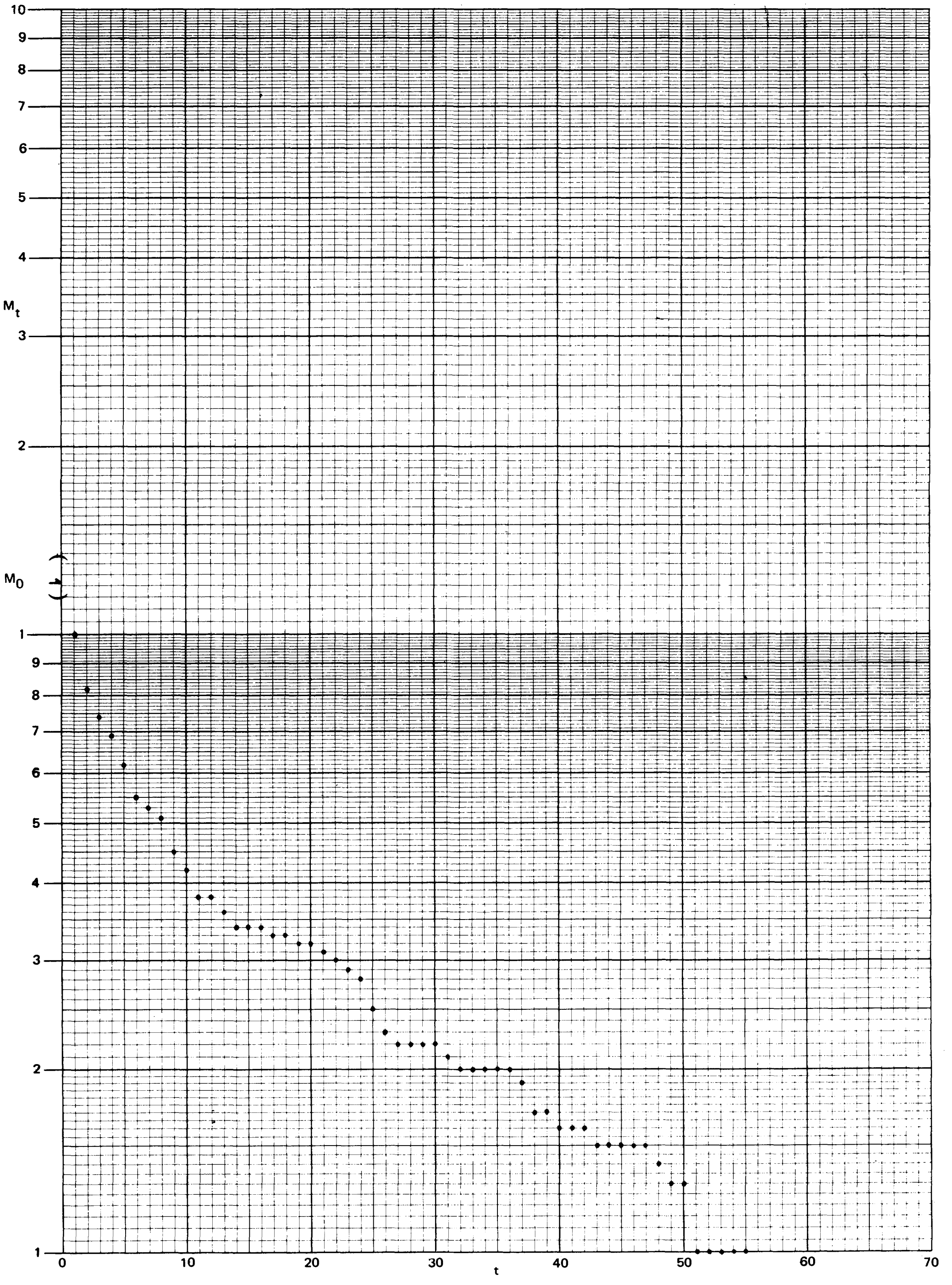

Fig. 2. M_0 extrapolation.

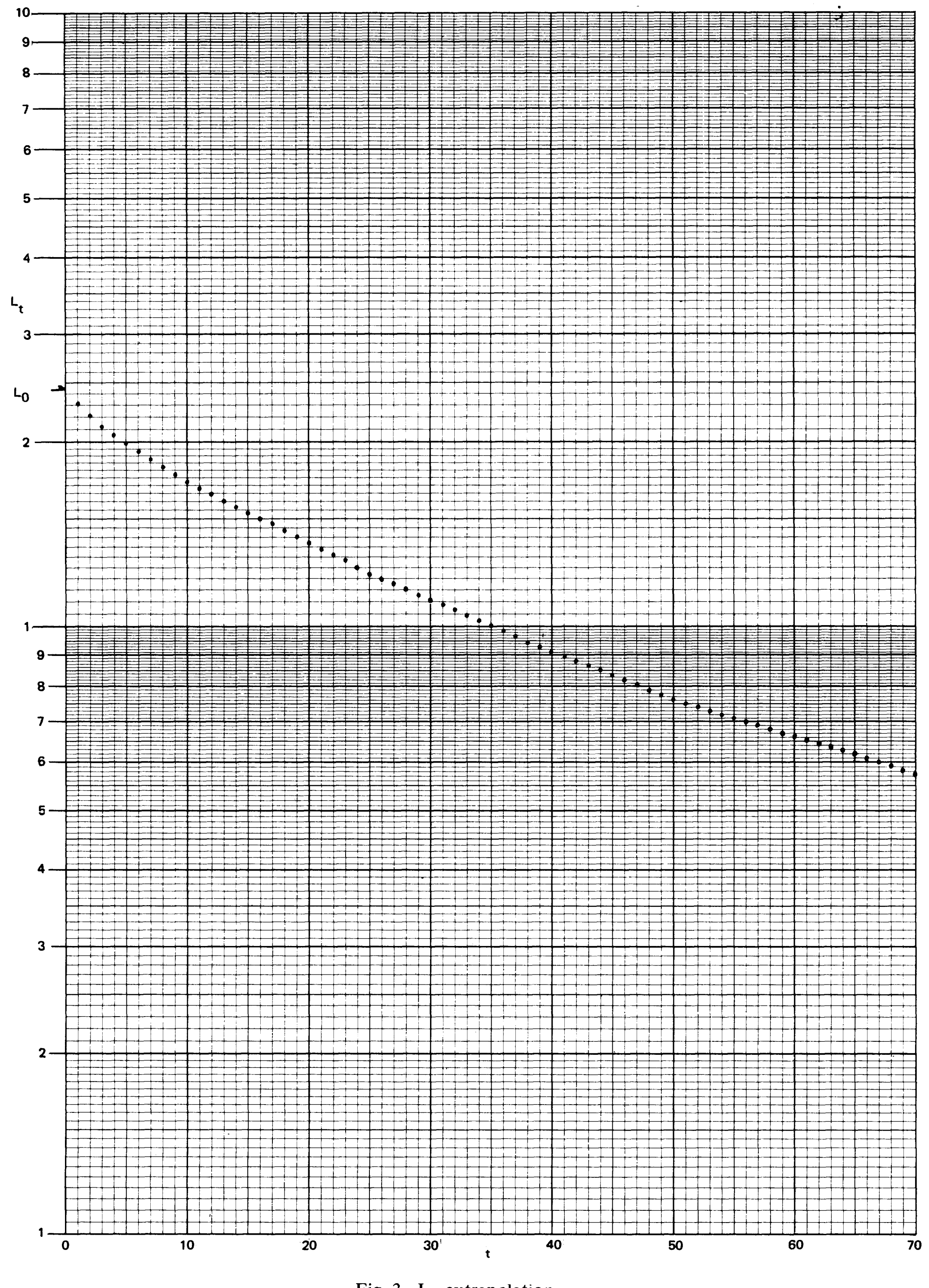

Fig. 3. L_0 extrapolation.

parabolic relationship may be seen to obtain in Figs. 2 and 3; however, fitting the data to a least-squares parabola did not produce a result which projected an M_0 as high as M_1, a necessary restriction. Least-squares was therefore rejected since it ignores restrictions on variable domains. A reasonable graphic extrapolation could result in a value between 115 and 133, with 122 as the best graphic extrapolation. In an independent verification of this range of values for the Joint site, Carole Wiley found 135 design elements occurring on sherds collected by surface sampling and excavation throughout the Hay Hollow Valley (Wiley 1971:5-6).

Finally, we note that a quick estimate for M_0, but one which is based upon assumptions of linearity between log M_t and t may be obtained by the formula $M_0 = (L_1{}^2/L_1 - M_1) - L_1$. This should never be used as a substitute for graphic extrapolation from the complete data tabulations, however, but only as an interim estimate. In the present case this method produces an estimate of 104 design elements.

Next, the frequency of occurrence of retrieved design elements was graphed against log r, where r is the rank by decreasing frequency of the design element and p(r) is the frequency of the element of rank r. The frequencies of occurrence of the extrapolated design elements was extrapolated based on the retrieved distribution and with $p(M_0 + 1) = 0$. This graph was also basically parabolic in shape, and a parabolic graph drawn through this data and intersecting the actual points at $r = 1$ and at $r = 123$ was used as the basis of the extrapolation. The graph paper used was about 11 ft long, that scale allowing an extrapolation to five decimal places. The result is summarized in Fig. 4, where the solid line indicates the extrapolated values and the points are those of the retrieved data.

The frequencies actually extrapolated were somewhat different from those indicated in Fig. 4 since, as was to be expected, they did not sum to 1. Since elements having occurred comparatively often can be expected to have stabilized about their actual frequencies more than those having occurred only a few times, it was decided to use a progressive scaling which would minimize deviations among the higher-frequency elements. The scaling progression was a linearly decreasing weighting system, which preserved the basic parabolic shape of the extrapolated rank-frequency curve.

Noise Levels

For design elements on ceramics, the presence of noise can only be manifested in the loss of transmitted elements since we do not expect to find different design elements in any instance than those that were applied by the prehistoric painter. To calculate the noise factor for each design element, let

$$\begin{aligned}\psi(r) &= P(\text{receiving design element r given that r was sent})\\ &= P(\text{receiving r} \mid \text{r was sent})\end{aligned}$$

But $P(A \mid B) = P(A \text{ and } B)/P(B)$, so

$$\begin{aligned}&P(\text{r sent} \mid \text{r received}) \cdot P(\text{r received})\\ &= P(\text{r sent and r received})\\ &= P(\text{r received} \mid \text{r sent}) \cdot P(\text{r sent})\end{aligned}$$

But $P(\text{r received} \mid \text{r sent}) = \psi(r)$; $P(\text{r sent}) = p_E(r)$, the extrapolated frequency for the design element r; $P(\text{r sent} \mid \text{r received}) = 1$; and $P(\text{r received}) = p(r)$. Thus, $\psi(r) = p(r)/p_E(r)$. To find the parameter λ of our code we must find the minimum of the $\psi(r)$ values, which in this case turned out to be 0.5059. But

$$\begin{aligned}\min_r \psi(r) &= \min_r P(\text{r received} \mid \text{r sent})\\ &= \min_r 1 - P(\text{r not received} \mid \text{r sent})\\ &= 1 - \max_r P(\text{r not received} \mid \text{r sent})\\ &= 1 - \lambda\end{aligned}$$

so that in this example, $\lambda = 0.4941$. This is low precision.

In calculating $\psi(r)$, any values exceeding 1 were recorded as 1 since it is a probability and therefore cannot exceed 1. Variation from the expected values along with the fact that the sum of the observed frequencies must sum to 1 in spite of being an incomplete sample can and will produce such discrepancies in the data.

One ambiguity left unresolved for the time being arose from a comparison of the observed and extrapolated frequencies, which showed that many of the values were not particularly close. This was confirmed by a computation of the expected variance in the occurrence frequencies, which for most elements revealed deviations far too great to be accounted the result of chance variation within the same population. The ambiguity is whether this is the result of a failure of the extrapolation technique due to a highly improbably skewed initial frequency distribution of design elements induced by differential preservation or whether it is a result of the noise levels being roughly as indicated from the results of that extrapolation. The ψ-distribution is plotted in Fig. 5, and the pattern is similar to that of the normal distribution, but one-tailed since 100% is the theoretical maximum. What we seem to have is random deviation from the mean value, which is computed to be 95.67%. This is supportive of, though it in no way proves, the appropriateness of the computed values. Against it is the fact that the observed values should never be higher than the extrapolated, and that the mean value should be 1, since

$$\bar{\psi} = \sum_r p_E(r)\,\psi(r) = \sum_r p_E(r)\left[p(r)/p_E(r)\right] = \sum_r p(r) = 1$$

The standard deviation of the ψ-distribution calculated is 0.66%, so that the observed values are not within 2 standard deviations of the mean. These arguments lose some force, however, when the observations of the preceding paragraph are taken into account. For those reasons, and for the reasons outlined above in this paragraph—especially since the rejection of the calculated results would require the acceptance of a statistically very unlikely skewing hypothesis—it seems fairly safe to accept the results and regard the differences between the observed and extrapolated distributions as being accounted for by the noise factor until the contrary can be shown.

Existence of a Code

At this point, we return to the problem of the existence of a code for the transmission of signals, the signals in this case being the combinations of design elements on a pot. The problem of finding the channel capacity leads in our example to the problem of solving 122 non-linear equations in 122 unknowns, a problem unapproachable without the use of a computer. An upper bound of 6.6438 was obtained by Helgert's method. The value of $C - \epsilon$ must be less than 1 to be at all restrictive, since 2^n is the number of different combinations than can be gotten from n elements, so that there will always be at most 2^n code words. Since the value of ϵ must lie between 0 and C, our range of values for $N = 2^{n(C - \epsilon)}$ is not limited any further. There will, therefore, at least if $C > 1$, be a value of ϵ for which the actual value of N is achieved, so that a code does exist in this case. This was a rather convenient result, since the design element tabulations were made from sherds, which did not allow a determination of this empirical N.

Information Distortion

One measure that has significance for archaeological interpretation is the relative entropy of the observed to the expected attribute distribution. This may be expressed very simply as the ratio H/H_E, where H is the information content of the system as recovered while H_E is the information content of the system as it was presumably operative prehistorically. If these entropy values are over 5 the significance of the differences between these values can be measured by the χ^2 test. If the results indicate a significant difference in the values, a ratio less than 1 indicates that the loss of data has resulted in a decrease in information from the original time of transmission, so that any interpretation of the data, however valid in terms of the modes used to interpret them, will be a

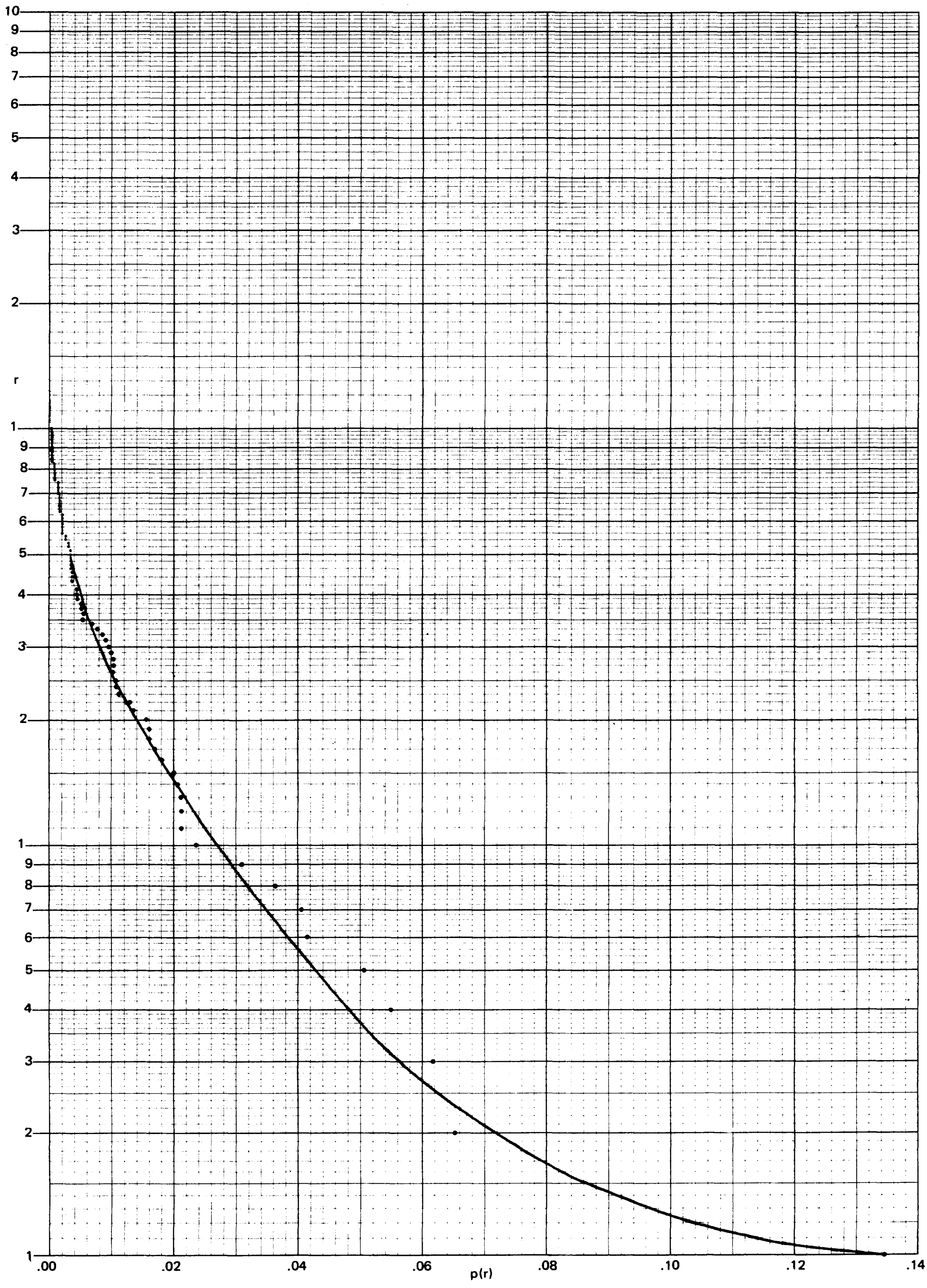

Fig. 4. Rank-frequency graph.

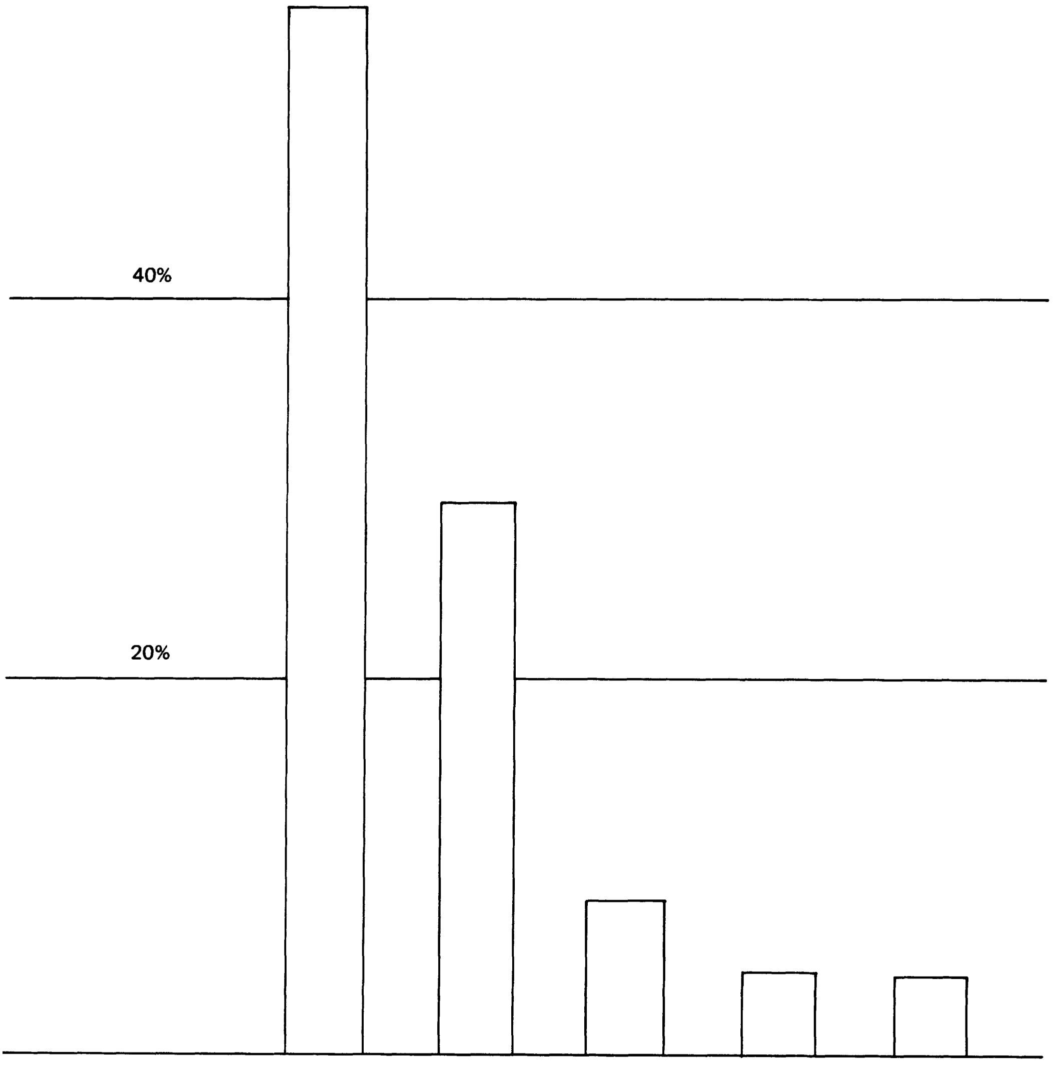

Fig. 5. ψ levels.

poorer reflection of the prehistoric situation by the percentage amount indicated by the ratio. A value greater than 1 is a measure of the degree to which the loss of data has created new patterns in the materials studied, the crucial data which would have nullified these patterns being unavailable. Simple algebraic manipulations show either of these situations to be possible. It should also be noted, however, that similar or even identical entropy values may be obtained from very different frequency distributions, so that an insignificant difference does not necessarily indicate a lack of significant information distortion.

The entropy values calculated from the present data were H = 5.3082 and H_E = 5.4180, giving a χ^2 value of .0022, which corresponds to the .02 level of significance. Thus, no significant distortion of information is *necessarily* indicated, though it cannot be ruled out. The distribution of noise over the design elements does seem normal enough that a low level of distortion is a reasonable interpretation in this case, however. This was perhaps to be expected considering that local conditions at the Joint site were highly favorable for preservation of a great variety of cultural remains, ceramics included. The ratio H/H_E was computed to be 0.9797. As an amount of loss, this is not significant, nor is it likely to be significant as distortion, as indicated above.

System Dynamics

Another measure that can be applied is simple relative entropy h, which is the proportion of "informativeness" in the system as opposed to $R = 1 - h$, the redundancy of the system. Redundancy results from the differential use of some attributes over others; more frequent use of a few elements over others gives less discrimination between alternatives than would be the case if all frequencies are equal. In fact, the maximum entropy value obtainable for a given number of attributes occurs when all frequencies are equal, giving, for k = total number of attributes

$$H' = -\sum_{i=1}^{k} p(x_i) \log_2 p(x_i) = -\sum_{i=1}^{k} \frac{1}{k} \log_2 \frac{1}{k} = -\log_2 \frac{1}{k} = \log_2 k$$

Thus, to obtain the simple relative entropy we calculate

$$h = H/H' \text{ and } h_E = H_E/H'_E.$$

Redundancy is a measure of the predictability of occurrence of elements due to chance variation, while the relative entropy h is the result of the operation of choice determining the use of attributes. There is very likely a mixing of both in any system of discriminating attributes so that redundancy may be thought of as a measure of the stability in the use of elements, relative entropy as the dynamics. In cultural systems, redundancy should be expected to be quite a bit less than relative entropy if cultural systems are thought of as essentially dynamic.

In our example, the values were calculated to be

$$h = 0.7990 \qquad R = 0.2010 \qquad h_E = 0.7956 \qquad R_E = 0.2044$$

The χ^2 test is not applicable to these data since the values are too low, but visual inspection alone suffices to show that the extrapolated values do not depart significantly from those for the retrieved data. We also find that design element use at the Joint site was the reflection of a relatively dynamic system. For comparison, the relative entropy of the Roman alphabet as used in standard written English is 0.836 and that of the Cyrillic for standard written Russian is 0.866 (Herdan 1966:278-279).

Binary Coding and Its Applications

Perhaps one of the most interesting applications of the entropy function results from the binary coding of the attributes being studied. A binary coding produces the theoretically most efficient

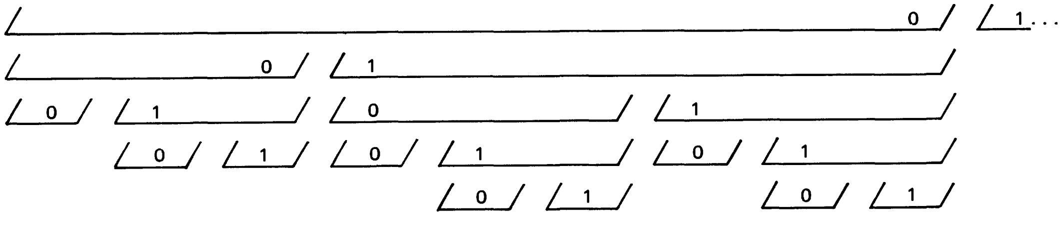

Fig. 6. Binary coding schematics.

Table 1. Binary coding of retrieved data.

r	Code	r	Code	r	Code
1	000	34	1101110	68	111110100
2	0010	35	1101111	69	1111101010
3	0011	36	1110000	70	1111101011
4	0100	37	11100010	71	111110110
5	01010	38	11100011	72	1111101110
6	01011	39	1110010	73	1111101111
7	0110	40	11100110	74	1111110000
8	01110	41	11100111	75	1111110001
9	01111	42	11101000	76	1111110010
10	10000	43	11101001	77	1111110011
11	10001	44	11101010	78	1111110100
12	10010	45	11101011	79	1111110101
13	100110	46	11101100	80	1111110110
14	100111	47	11101101	81	11111101110
15	101000	48	11101110	82	11111101111
16	101001	49	111011110	83	11111110000
17	101010	50	111011111	84	11111110001
18	101011	51	11110000	85	11111110010
19	101100	52	11110001	86	11111110011
20	101101	53	11110010	87	11111110100
21	101110	54	111100110	88	11111110101
22	101111	55	111100111	89	11111110110
23	110000	56	111101000	90	111111101110
24	1100010	57	111101001	91	111111101111
25	1100011	58	111101010	92	11111111000
26	110010	59	111101011	93	11111111001
27	1100110	60	111101100	94	11111111010
28	1100111	61	111101101	95	11111111011
29	110100	62	111101110	96	11111111100
30	1101010	63	111101111	97	11111111101
31	1101011	64	111110000	98	11111111110
32	1101100	65	111110001	99	111111111110
33	1101101	66	111110010	100	111111111111
		67	111110011		

coding possible for any set of data since it discriminates by binary choices into equal frequency categories. To set up the binary code for an attribute system, the attributes should first be ranked by frequency from highest to lowest; then the frequencies are divided into 2 groups with equal frequency totals, or with totals as nearly equal as is possible. The first group receives the code digit 0, and is made up of the higher-frequency elements, the second receives the digit 1 and is made up of the lower-frequency elements. Then the process is applied to each of these subgroups, and then continually to the resulting subgroups until all the attributes have been isolated in single-attribute groups. This process is represented schematically by Fig. 6. The binary codings themselves are given in Tables 1 and 2.

Given this data we compute H*, the information content of the system in terms of binary coding, or, in other words, the maximum information retrievable for a given frequency distribution, by the formula

$$H^* = \sum_{i=1}^{k} p(x_i)\, b(x_i)$$

where $b(x_i)$ is the number of digits in the binary code for attribute x_i, and the other quantities in the formula are as in the last section. The ration $h^* = H/H^*$ is then a measure of the coding efficiency.

The degree of efficiency in terms of binary coding is a measure of the relative degree of discrimination within the total system of design elements. The choice of attributes used in any system is assumed to reflect choices between the aesthetic, ideological, representational or functional ideas with which these elements are consciously or unconsciously associated by the people who made these choices or by the people for whom they were made. We should expect a high efficiency value, reflecting a high level of discrimination of ideas, in any cultural subsystem. A low degree of efficiency characterizing the use of elements of an attribute system would reflect poor discrimination of ideas. In this case, either these ideas or governing principles were so diffuse or confused in the minds of their makers as to be incoherent and unintelligible, perhaps even inconsistent with each other; or they reflect an inappropriate degree or form of discrimination on the part of the archaeologist. As mentioned above, the first possibility seems unlikely except perhaps in a phase of cultural transition or a phase immediately preceding one. Otherwise, we should look to the sources of confusion in discrimination resulting from the analysis. (1) The units of analysis may be inappropriate; that is, the attributes which the archaeologist discriminates within the system may be poor reflections of those discriminated by the prehistoric population. Alternatively, there may be differences in the way attributes discriminated are used and the ideas they reflect for (2) different groups of people within the site, (3) differences across time, (4) contextual differences, (5) functional differences, or (6) any combination of these.

Moreover, if significant differences did exist in attribute discrimination based on any of the above-mentioned factors, we would then expect a lumping to result in a decrease in overall discrimination because of discrimination confusion. Thus we should have no such confusion if we have high efficiency—over 95%, for example. For our data the values are $h^* = 99.21\%$ and $h^*_E = 97.12\%$; here we do have very high efficiency, so that we should expect a fairly high degree of homogeneity in the cognitive discriminations reflected in ceramic design elements across temporal, spatial, social, contextual, and functional units at the Joint site.

These results also support the design element discrimination used as the basis of this study and the validity of design elements rather than necessarily some higher taxonomic grouping of them as good reflections of prehistoric ideational sets. Had the typology used not been so successful the elements could have been rediscriminated or else grouped into larger non-overlapping taxonomic units; the recalculation of h* for each typology gives a basis for choice between them and for an evaluation of them. Most reasonable typologies—typologies reflecting the same cognitive sets in alternate ways—should apparently give good results. Four such typologies were tested in which classification was either by shape, by discriminating features within shape, or by combinations of the two. One of these was based on a deliberately somewhat cruder scheme. The h* values for the 3 most reasonable typologies of the design elements were 99.14%, 99.41%, and 99.46%, while the somewhat cruder typology had a lower value, 93.58%. A very unreasonable typology with no apparent internal relations used to classify the elements into its groups gave an efficiency value of only 0.16%, so that the above results are not the inherent consequents of any regrouping of elements but are specific to the regroupings used. If these results are generalizable, then it means that human beings make multiple discriminations of phenomena in the same system, reflecting concurrent dimensions for structuring what they perceive as a single cognitive domain. This is in

Table 2. Binary coding of extrapolated data.

r_E	Code	r_E	Code	r_E	Code
1	000	42	1110010	82	1111110001
2	0010	43	11100110	83	1111110010
3	0011	44	11100111	84	1111110011
4	0100	45	11101000	85	1111110100
5	01010	46	11101001	86	11111101010
6	01011	47	11101010	87	11111101011
7	0110	48	11101011	88	1111110110
8	01110	49	11101100	89	11111101110
9	01111	50	11101101	90	11111101111
10	10000	51	11101110	91	11111110000
11	100010	52	111011110	92	11111110001
12	100011	53	111011111	93	11111110010
13	10010	54	11110000	94	11111110011
14	100110	55	111100010	95	11111110100
15	100111	56	111100011	96	11111110101
16	101000	57	11110010	97	11111110110
17	101001	58	111100110	98	11111110111
18	101010	59	111100111	99	11111111000
19	101011	60	111101000	100	111111110010
20	101100	61	111101001	101	111111110011
21	101101	62	111101010	102	111111110100
22	101110	63	111101011	103	111111110101
23	101111	64	111101100	104	111111110110
24	110000	65	111101101	105	111111110111
25	1100010	66	111101110	106	111111111000
26	1100011	67	111101111	107	111111111001
27	110010	68	111110000	108	111111111010
28	1100110	69	1111100010	109	1111111110110
29	1100111	70	1111100011	110	1111111110111
30	1101000	71	111110010	111	111111111100
31	1101001	72	1111100110	112	1111111111010
32	1101010	73	1111100111	113	1111111111011
33	1101011	74	111110100	114	1111111111100
34	1101100	75	1111101010	115	11111111111010
35	1101101	76	1111101011	116	11111111111011
36	1101110	77	1111101100	117	11111111111100
37	11011110	78	1111101101	118	11111111111101
38	11011111	79	1111101110	119	11111111111110
39	1110000	80	1111101111	120	111111111111110
40	11100010	81	1111110000	121	1111111111111110
41	11100011			122	1111111111111111

partial verification of the componential analysis approach of ethnoscience, and the use of binary coding may nullify Burling's criticism of the approach as "arbitrary" (Burling 1966) in that it can verify cognitively meaningful distinctions.

Summary

In Part II several tests have been presented for dealing with data, and various findings and suggestions have been made. We have methods to extrapolate the number of elements and their frequency distributions in prehistoric attribute systems; to estimate noise levels; to determine the existence of an archaeological code for attribute systems; to measure the significance of distortion of information; to measure system dynamics and levels of cognitive discrimination reflected in the system; and to choose the "best" typology for an attribute system. Hypotheses requiring further investigation are that all cultural systems are essentially dynamic; that all cultural systems have high levels of attribute discrimination; and that any cognitive system uses multiple classification of phenomena in its domain. We have found the Joint site to conform to these hypotheses. In addition, we have suggested uniformity of cognitive subsystems within the site measured along various dimensions where variation might be expectable, and have found very low distortion of information within the material system considered.

CONCLUSION

In addition to providing the theoretical justification for the methodological applications of Part II, most of the discussion of Part I was intended as a theoretical framework for a research design then being formed which it was hoped would result in some conclusive generalizations on the limits of our inference beyond simply the ascertainment of these limits in any particular instance. It turns out that general considerations provide no bound on the interpretability of archaeological data.

It should be cautioned that the central concepts of channel and code are only a model relating a set of input variables to a set of output variables; thus, information theory is only one specialized method of statistical inference, many others of which might be more helpful in reconstructing certain facets of the physical distributions of material artifacts than would an information-theoretic approach. The special usefulness of information theory enters where we have cultural systems that can be treated as communication systems, or where the anthropological studies are themselves scrutinized as they articulate with the cultural system under investigation.

The remainder of the article was an attempt to show the value of the entropy function as a tool in archaeological research, almost to the point of being the basic element of one coherent approach. It is by no means an attempt to present this approach as *the* way to approach all archaeological problems; rather, it is an exposition of an ancillary method of attacking problems and forming and testing hypotheses, and of estimating the reliability of the data for inferential purposes. This last use, as applied here, was only to a specific site, so that use of comparative material when relevant may considerably expand the usefulness of the data, in which case the limitations would have to be studied in conjunction with this material as well.

Acknowledgments. Thanks are due the Undergraduate Research Participation Program of the National Science Foundation, Grant GY-8938, and to the Field Museum of Natural History for making it possible for me to acquire the data for this paper. Frederick Gorman and Ezra Zubrow of the Southwest Archaeological Expedition made very direct contributions to the germination and progress of this paper. David Blackwell of the Department of Statistics at the University of California, Berkeley, read the rough draft and offered suggestions which I gratefully accepted.

Burling, Robbins
1966 Cognition and componential analysis: God's truth or hocus-pocus? *American Anthropologist* 66:20-28.

Gorman, Frederick
1970 Symbol as strategy: toward a synthetic view of material culture. Department of Anthropology, Field Museum of Natural History. xeroxed.
Hatch, Marion P.
1971 An hypothesis on Olmec astronomy, with special reference to the La Venta site. In Papers on Olmec and Maya Archaeology, no editor. *Contributions of the University of California Archaeological Research Facility* 13:1-64.
Helgert, Hermann J.
1967 On a bound for channel capacity. *IEEE Transactions on Information Theory* 13:124-126.
Herdan, Gustav
1966 *The advanced theory of language as choice and chance.* Springer-Verlag, New York.
King, Thomas F.
1971 A conflict of values in American archaeology. *American Antiquity* 36:255-262.
Kotz, S.
1966 Recent results in information theory. *Methuen's Review Series in Applied Probability, Volume* 5.
Mackay, Alan
1965 On the type-fount of the Phaistos Disk. *Statistical Methods in Linguistics* 4:15-25.
Margalef, Ramon
1968 *Perspectives in ecological theory*. University of Chicago Press, Chicago.
Postal, Paul M.
1971 On the surface verb "remind." In *Studies in linguistic semantics*, edited by Charles J. Fillmore and D. Terence Langendoen, pp. 180-270. Holt, Rinehart, and Winston, New York.
Quastler, Henry
1956 Studies of human channel capacity. In *Information theory; third London symposium*, edited by Colin Cherry, pp. 361-371. Academic Press, New York.
Sonntag, Richard, and Gordon Van Wylen
1966 *Fundamentals of statistical thermodynamics*. Wiley, New York.
Wiley, Carole
1971 Social interaction and economic exchange in the Hay Hollow Valley 900-1200 A.D. Department of Anthropology, Field Museum of Natural History. mineographed.
Wolfowitz, J.
1961 *Coding theorems of information theory*. Springer-Verlag, Berlin.