I

The diversity concept

Chapter 1

The concept of diversity: an introduction

George T. Jones and Robert D. Leonard

The adoption of an evolutionary perspective on cultural change with an emphasis on empirical variability, its transmission and differential representation through time, marks a significant trend in modern archaeology (Dunnell 1980). It is quite common for archaeologists to examine how the archaeological record differs along gradients of various sorts, through time, for example, or across space. In this context, archaeologists devote considerable energy to describing and attempting to explain the patterns perceived. Virtually every model employed to achieve these ends has definite entailments for the distribution of artifacts among different classes on such gradients. This is the basic matter of diversity: how quantities of artifacts are distributed among classes.

While the adoption of an evolutionary perspective which focuses on variation and incorporates the concept of diversity in a rigorous manner is a relatively recent phenomenon, observations regarding variety, and thus diversity, have a long history. Indeed, such notions are fundamental, being basic to any discipline where phenomena are arrayed into a number of classes, and where those classes can have differing numbers of members. Of course, this is a common archaeological situation, and recognition of the great variety of material remains in archaeological settings was known in American archaeology well before the turn of the century. By 1919, the variety of material remains across the Americas was well enough characterized for W. H. Holmes (1919) to generalize about issues related to diversity, as regards a number of culture areas. Consider, for example, the following passages, beginning with a discussion of the Middle and Lower Mississippi Valley area (Holmes 1919: 107):

> As a result of the mineral riches of the area, the range of lithic artifacts is greater than in any other region north of the Valley of Mexico.

The North Andean–Pacific area (*ibid*: 135):

> It is noteworthy . . . that much diversity [regarding antiquities] is shown, especially in the more southern districts; a condition due partly, it would appear, to intrusive elements of race and culture as well as to isolation of communities by reason of the pronounced physical characteristics of the country.

The Middle Andean–Pacific area (*ibid*: 137):

> In no other area do the antiquities have so wide a range or tell so completely the story of the life and fate of the people. Above ground the more durable artifacts and works only are preserved, but in the depositories of the dead, especially in the arid districts, vast numbers of every conceivable product of handicraft are preserved almost unchanged.

And Primitive [sic] South America (*ibid*: 143):

> Stone implements and utensils of excellent make and considerable variety are widely distributed.

These passages exemplify that notions of diversity have had a long, albeit intuitive, formulation in archaeology. Likewise, it is apparent that archaeologists (Holmes at least) sought to characterize and explain diversity as a consequence of social and environmental conditions. It is through quantification that notions of diversity lost, for the most part, their intuitive component. The notions became measurable and, as such, gained access to assessments of validity and reliability. The passages also illustrate that issues of diversity are of concern in virtually all archaeological settings, although it often goes unrecognized even today.

Although diversity is a quite simple concept, there is some confusion surrounding its meaning. In part, this confusion arises from a rather substantial terminology which we will try to make clear, and through the concept's implicit beginnings. Moreover, the term 'diversity' has been employed as a synonym for variation, which it is not; diversity is a measure of variation. Specifically, diversity refers to the nature or degree of apportionment of a quantity to a set of well-defined categories (Patil and Taillie 1982). Thus, diversity is a referent for the structural properties of a population or sample made up of distinct categories. We can think of diversity as an average property, that is, a measure of variation composed as a single value.

Generally, diversity is rendered either as the number of categories represented in a sample or as the manner in which a quantity is distributed among those categories. These concepts are respectively termed richness (the number of classes), and evenness (the order of abundance values). Some confusion surrounds the fact that a number of familiar indices of diversity combine these components in a single measure. Peet (1975, following Good [1953]) has chosen to term this dual concept of diversity as heterogeneity. However, others like Pielou (1975) refer to such statistics, which confound the number of classes and evenness, simply as diversity. An attempt to quantify evenness apart from richness is embodied in the notion of equitability. Measures of equitability assess evenness relative to some standard. Commonly, that standard is a theoretical distribution like a geometric or logseries distribution or a maximum possible value of heterogeneity for a given sample size and class number.

It is important to remember that indices of diversity constitute a powerful set of tools for characterizing the structure of populations and samples. When the measures employed to examine structure confound richness and evenness, it is always necessary at some level to break these measures down into the constituent components, to search for what may be a more parsimonious characterization. This is necessary because, unknown to the investigator, one or the other component (richness or evenness) may be the primary determinant of the value of the composite measure. Moreover, when two or more samples are compared, values of the composite measure across samples may be identical, yet the structure of richness and evenness may be extremely different. Such a characterization, while accurate, might lead to naive, or even erroneous, conclusions. It must also be considered an error when composite diversity indices formulated as information statistics (e.g., Shannon and Weaver 1949) are applied to values obtained as measures of information, rather than characterizations of population or sample structure, without the direct correspondence between artifact and information being known.

We do not want to give the mistaken impression that the concept of diversity and its many measures are equivalent, although the distinction often is not clear in the literature. As we shall see in the following papers, not all measures of diversity supply consistent information; all measures are not equally sensitive to different portions of class abundance curves. This does not mean that we should be pessimistic about the concept of diversity as a useful construct, because without the measure of diversity we actually have very little knowledge of our materials. This only means that we need to exercise selectivity in our choice of measure. Because there are these inconsistencies among measures, Patil and Taillie (1982) have suggested that we recognize an intrinsic diversity ordering that is independent of the indices in use. Such an order meets two criteria. First, with respect to richness, we say that the sample with the greatest number of classes represented is most diverse. Stated more formally, richness is increased as classes are successively added to a sample. Secondly, a maximally even distribution is most diverse. That is, diversity increases when abundance is transferred from one class to another, less abundant, class. Thus, we say that sample A is intrinsically more diverse than sample B if it contains more classes and has a more even distribution of class members.

In practice, the actual orders generated by diversity indices may depart from the intrinsic diversity order because samples are not necessarily comparable. This brings us to a number of requirements that must be met in order that comparisons of values of diversity may be made. Although these are detailed more fully in later papers, we will mention several here. First, the measurement of diversity rests on an unambiguous classification of the subject matter. This means that classes must be defined that are mutually exclusive, exhaustive, and composed at the same classificatory level. Secondly, if samples are being measured, they must be generated either randomly, or in some other manner be representative with respect to the gradients over which differences are evaluated. This means, for instance, that if the diversity of artifact samples from different microenvironments is compared, the samples must be a true reflection of the populations of artifacts from each zone. Because diversity, specifically richness, assumes a dependent relationship with sample size, several indices have been proposed as independent. Such indices make the further assumption that the relationship between sample size and richness remains constant over the samples compared. Other, measure-specific, assumptions also exist.

The reasons why archaeologists might be concerned with the measurement of diversity seem apparent enough, but let us discuss two general points. First, the structure of most archaeological data is such that it invites quantitative description. Diversity constitutes a measure of our perceptions of those data. The mathematical challenge in representing that perception as a single statistic has certainly led in part to the interest in diversity indices. We do not suggest that this is the sole reason or the most important reason for investigating diversity though it does account for the rather weighty mathematical underpinnings of the concept (e.g., Pielou 1975; Patil and Taillie 1982). For our purposes, concerns with the quantitative properties of diversity are rather sterile if not accompanied by an equal interest in diversity as a property of archaeological phenomena. Thus, we add a second reason for interest, which is that, by measuring archaeological diversity, we may provide the means to examine the nature of processes that govern the representation of different classes of phenomena in the archaeological record. Diversity studies then offer potential for resolving functional and processual relationships.

The authors of the following chapters take up both of these issues, variously exploring the statistical properties of diversity with respect to archaeological data and the place of diversity in models of archaeological explanation. We can all imagine that some of the inherent qualities of the archaeological record make study of the former issue critical, given the assumptions of diversity measures. It may be that, in coming to grips with diversity measured in an archaeological setting, we can offer a unique perspective on the general properties of diversity. On one point, the relationship between sample size and diversity, a number of chapters make such a contribution. This comes about because archaeologists deal with a very different set of phenomena than do biologists, in whose discipline most of the measures were developed. In biological applications, rarely can entire populations be censused but, most often, extreme precision regarding the context and size of samples taken can be maintained. In archaeological settings, however, we may deal with either complete populations, or samples of discrepant sizes. This situation is often, although not always, beyond the control of the investigator. Even in those instances where we have little doubt that we have a set of data that constitutes a population, this population may also be considered a sample in the context of human behaviour within a given time period and geographical expanse. Much of the potential of our contribution lies in addressing these issues.

While the following chapters address a large number of topics relating to diversity, we have by no means touched upon, let alone exhausted, the range of topics worthy of discussion. One issue only alluded to in the following pages, and of certain interest, centers on the classificatory units we employ, whether they are real or contrived, and how we should go about constructing them. Surely this point must bear on the meaning we assign to values of diversity.

We have just begun to explore the problem contexts in which this quantitative description of variation assumes significance. The research presented here joins a small, but growing, body of literature on the study of archaeological diversity. We think that it is safe to say that we are now at the stage of evaluating the progress of diversity as a useful archaeological construct.

Chapter 2

The theory and mechanics of ecological diversity in archaeology

Peter T. Bobrowsky and Bruce F. Ball

A perusal of recent literature indicates that *diversity* has become a popular concept in archaeological research, but observations like that made by Reid (1978:203) that '*Diversity* is what the diversity index measures' suggest a basic misunderstanding of the concept of diversity. There exist important differences in the application of the concept of diversity in archaeology, so that a formal examination of the concept is needed.

Perhaps the most elementary differentiation exists in its use in a qualitative sense (e.g., Binford 1982; Hayden 1981; Schiffer 1983) as opposed to its specific meaning in a quantitative analysis (e.g., Cannon 1983; Grayson 1984; Jones *et al.* 1983; Kintigh 1984a). Diversity is a well-defined and widely used concept in ecology. A variety of equations exists in ecological research to measure diversity; this is no less so in archaeology. Applications of the concept of diversity in archaeology, as well as the equations designed to measure diversity, abound. If everyone is measuring diversity with a different equation, either the term has multiple connotations or all equations are equivalent. The purpose of this study is to illustrate the complexity inherent in the concept of diversity and to demonstrate why a singular definition and measure is inadequate, misleading, and leads to inaccurate interpretation. In the following discussion we (1) reduce the concept of diversity into separate components known as richness, evenness, and heterogeneity; (2) examine the ecological characteristics of the three components; (3) provide and decompose equations which measure the characteristics of the components; (4) provide a historical review of the use of the concept in archaeology; and (5) examine the archaeological examples with regard to the formal ecological definitions.

Quantitative archaeology subsumes a broad range of activities, thus allowing archaeologists to count, measure, and weigh artifactual remains in a variety of ways. Nonetheless, one archaeologist recently concluded that the quantification and comparison of artifactual assemblages can be simplified to three basic forms. According to Cannon (1983:785), these three basic quantitative measures are: (1) absolute counts; (2) proportional frequency; and (3) diversity. All other measures may be interpreted as permutations and derivatives of the above. Following a review of the three types, Cannon concludes that diversity is the only reliable measure. This conclusion in itself warrants a detailed examination of diversity measures. Moreover, it can be shown that by definition both counts and proportions define aspects of diversity. If Cannon's thesis remains acceptable after our clarification, then all measures are in some manner related to diversity. In the discussion that follows, we support assertions and equations surrounding diversity with appropriate citation. However, we do not intend to provide an exhaustive review of the literature. Instead, we direct the interested to a treatise by Grassle *et al.* (1979) which contains over one thousand references on the subject of diversity.

Ecological concept of diversity

As generally understood, diversity describes complex interspecific interactions between and within communities under a variety of environmental conditions. By necessity, diversity description requires quantitative descriptors besides simple qualitative appraisal. As a natural consequence of this process, one witnesses the continued generation of new equations, each purportedly measuring diversity. Most of these new equations are simply modifications of existing equations; unfortunately they are also usually more complex, unmanageable, and probably unnecessary. Before adopting one or several such equations to describe diversity, one requires a clear understanding of what aspect of the interspecific interactions is actually being measured. In short, we suggest that a detailed decomposition of the concept and its associated viable measures will prove productive for archaeologists interested in its use.

The general concept of diversity embodies three distinct aspects or components: (1) richness; (2) evenness; and (3) heterogeneity. Following Hurlbert (1971:581), numerical *species richness* is defined as 'the number of species present in a collection containing a specified number of individuals'. Thus, in the biological usage, richness designates the variety of taxa, species, or types in an assemblage or community. Closely allied to species richness is areal species richness or *species density* which denotes 'the number of species present in a given area or volume of the environment' (Hurlbert 1971:581). Although allied, richness and density require distinction when quantitative appraisal is being attempted. In contrast, *species evenness* is considered to represent the absolute distribution of individuals across all species. Evenness attempts to describe the similarity in abundance of several species in the community. When evenness is compared to some given theoretical distribution, the resultant descriptors delimit *species equitability* (Lloyd and Ghelardi 1964). An unacceptable synonym of equitability is *relative diversity* (Sheldon 1969). The final component of diversity, termed *heterogeneity* (Peet 1974), reflects a dual concept in which richness and evenness are simultaneously measured. Heterogeneity is a measure that assesses the variability in both the numbers of species and the abundance of individual species with a single value.

Given the above definitions, general use of the term diversity is therefore equivocal unless distinctions are made as to which of the underlying characteristics is of concern. All three aspects uniquely describe different properties of community structure. Casual discussion and measurement of the properties of a community or assemblage invalidates resultant interpretations, whether ecological or archaeological. We suggest that the use of the general term *diversity* be abandoned in archaeological studies and be replaced with the threefold concept defined by ecologists.

Richness

The wealth or variety of species in a collection of individuals – *richness* – is an accessible property and provides a means by which differences or similarities in collections can be measured and compared. One of the simplest measures of richness favored by several researchers (e.g., MacArthur 1965; Williamson 1973) is the direct species count. With this approach, most community or population characteristics are ignored, and the observed variety of species in two or more assemblages is simply compared. A simple comparison of species counts ignores varying sample sizes. Since it is known that species richness is functionally dependent on sample size, simple species counts may be problematic. As evident in species area curves (Gleason 1922, 1925) and species individual curves (Odum *et al.* 1960), the number of types of species encountered in a collection increases asymptotically as the total area or number of individuals recovered increases. This asymptotic relationship and functional dependence has been confirmed by collector's curves for pollen (Duffield and King 1979), gastropods (Bobrowsky 1983) and benthic fauna (Sanders 1968). To circumvent this problem of sample-size dependence, one can compare collections containing equal numbers of individuals, or unequal samples of completely inventoried populations (see also Jones *et al.* 1983; Kintigh 1984a; and Grayson, Leonard *et al.*, Kintigh, and Schiffer in this volume in regards to sample-size considerations in archaeology). In practice it is difficult to obtain samples with the same number of individuals. Moreover, how many individuals constitute an adequate sample size? Palynologists prefer large but frequently unequal samples of 250 or more grains, thus assuming universal threshold points exist.

A more realistic alternative to the above dilemma is to assume that the relationship between species and individuals, or types and specimens, is constant and quantifiable within communities. The assumption appears valid as shown by the following measures of species richness which numerically describe a quantifiable relationship between species and individuals/area:

$$R_1 = (S - 1)/1nN \qquad \text{(Margalef 1958)} \qquad (1)$$

$$R_2 = S/\log N \qquad \text{(Odum et al. 1960)} \qquad (2)$$

$$R_3 = S/\sqrt{N} \qquad \text{(Menhinick 1964)} \qquad (3)$$

$$R_4 = S/\log A \qquad \text{(Gleason 1922)} \qquad (4)$$

$$\hat{S}_1 = \alpha 1n(1 + N/\alpha) \qquad \text{(Fisher et al. 1943)} \qquad (5)$$

$$\hat{S}_2 = y_0\tilde{\sigma}(2\pi)^{1/2} \qquad \text{(Preston 1948)} \qquad (6)$$

$$\hat{S}_3 = 2.07(N/m)^{0.262} \qquad \text{(Preston 1962)} \qquad (7)$$

$$\hat{S}_4 = 2.07(N/m)^{0.262}A^{0.262} \qquad \text{(MacArthur 1965)} \qquad (8)$$

$$\hat{S}_5 = kA^{d} \qquad \text{(Kilburn 1966)} \qquad (9)$$

$$\hat{S}_6 = aN/(1 + bN) \qquad \text{(de Caprariis et al. 1976)} \qquad (10)$$

$$\hat{S}_7 = \sum_{i=1}^{s}\left[1 - \frac{\binom{N - N_i}{n}}{\binom{N}{n}}\right] \qquad \text{(Hurlbert 1971)} \qquad (11)$$

where:

S = the number of observed species,
N = the number of individuals in a collection,

A = the area of the isolate or collection,
m = the number of individuals in the rarest species,
α = Fisher's slope constant,
y_0 = the number of species in the modal class interval,
$\tilde{\sigma}$ = the estimate of the standard deviation,
n = the number of individuals in a subsample,
N_i = the number of individuals in the ith species,
R = the constant of rate increment,
$\hat{S}$ = the number of expected or predicted species, and
k, d, a and b = empirically derived coefficients of regression.

All of the above equations have been reviewed previously (e.g., Buzas 1979; Fager 1972; May 1975; Peet 1974; Whittaker 1972). Given the results of research in mathematical ecology and the goals of archaeological quantification, we limit our discussion to three reliable measures of richness which may be profitably used by archaeologists.

Equation (6) describes what Preston (1948) calls the truncated lognormal distribution. This equation illustrates a method by which the number of species in a population can be determined from varyingly sized but randomly collected samples. In the derivation of the truncated lognormal distribution, one assumes that the distribution of the totally inventoried population is a complete lognormal distribution. The assumption is acceptable given the arguments of May (1981), who notes that most distributions, from GNP between nations to diatoms in a stream, are defined by lognormality. However, since most samples are incomplete representations of the sampled population, a portion of the lognormal distribution is absent, hence the truncated lognormal distribution. Application of the truncated lognormal distribution to various samples allows one to calculate the sample-specific degree of truncation (= the number of species missing in the sample but present in the population) and therefore permits the estimation of the total number of species (richness) in the population. Additionally, this distribution has a secondary advantage; namely, evenness can also be evaluated from the derived sample statistics (see later discussion).

The second measure of species richness we consider is that provided by de Caprariis and colleagues (1976, 1978, 1981). Their equation (10) is a rectangular hyperbola generated by simple regression of inversely transformed data. Following several of their algorithms, one notes that, not only can the maximum value of species richness be determined, but optimal sample sizes can be estimated for varying fractional deviations using the following:

$$\varepsilon = [1/(a/b)][(a/b) - an/(1 + bn)] \tag{12}$$

where ε is the expression of the fractional deviation at the limiting value a/b (ratio of regression coefficients) and sample size n. Having chosen an acceptable percentage error one solves for n as:

$$n = (1 - \varepsilon)/b\varepsilon \tag{13}$$

The final method of species richness estimation relates to the rarefaction technique of Sanders (1968). As originally proposed by Sanders, the technique is invalid (Fager 1972; Heck *et al.* 1975; Raup 1975; Simberloff 1972; Tipper 1979). Fortunately, a decomposition and proper revision of the method has been offered by Hurlbert (1971) and is provided as equation (11). A review of Hurlbert's family of measures by Smith and Grassle (1977) indicates that equation (11) applies to a finite population where sampling is without replacement. This expression is an adequate approximation of sampling from an infinite population as described by the following:

$$S(m) = \sum_{i=1}^{\kappa} [1 - (1 - \pi_i)^m] \tag{14}$$

where π_i is the proportion of individuals in the ith species in a sample of size m. Further, it can be shown that when m equals two in equation (14) the function is related to Simpson's (1949) measure:

$$\lambda = \sum_{i=1}^{s} \pi_i^2 \tag{15}$$

where π_i is the proportion of individuals in the ith species (further discussion is provided under the topic of heterogeneity). An added advantage to equation (11) is that $\hat{S}_7$ has an unbiased minimum variance estimator (Smith and Grassle 1977).

In summary, each of the above three equations (6, 10 and 11) provides an estimate of species richness for differing sample sizes. Additionally, the truncated lognormal distribution can be examined for species evenness, the rectangular hyperbola with modification permits optimal sample-size calculations, while Hurlbert's measure allows for a variance estimate. Choice of particular species richness equations should vary between researchers depending on the secondary information required (cf. Bobrowsky 1983; May 1975; Wolda 1983).

Evenness

Knowledge of the species richness is indispensable to the study of diversity, but in itself fails to provide insight into particular underlying abundance distributions. In the analysis of any assemblage, one also requires knowledge of the frequency of representation of the contributing species. The basic question, whether all species are equally abundant, or certain species are more abundant than others, characterizes the necessity for measuring evenness.

According to Hurlbert (1971) and Peet (1974), most measures of evenness and equitability fall into one of two classes as defined by the equations:

$$E' = \Delta/\Delta_{max} \tag{16}$$

$$E = (\Delta - \Delta_{min})/(\Delta_{max} - \Delta_{min}) \tag{17}$$

where:

Δ = the observed value of the diversity parameter,
Δ_{max} = the maximum value attainable by the diversity parameter, and
Δ_{min} = the minimum value attainable by the diversity parameter.

We note that Pielou's (1975, 1977) popular measures:

$$J = H/H_{max} \tag{18}$$

and

$$J' = H'/H'_{max} \quad (19)$$

conform to equation (16) when H is Brillouin's index and H' is the Shannon and Weaver index (see discussion of heterogeneity below). Again, Δ in equation (16) may be replaced with Simpson's λ or any other measure as shown by Sheldon (1969) and Peet (1975). All of the above substitutions, H, H', and λ, apply equally well to Δ in equation (17). Similarly, Margalef's (1958) redundancy measure of evenness, given as:

$$R = (\Delta_{max} - \Delta)/(\Delta_{max} - \Delta_{min}), \quad (20)$$

derives its formulation from equation (17).

The evenness measures discussed thus far suffer from several inherent limitations. Briefly, these restrictions include a dependence on sample size, the species richness, and most importantly, the particular measure used in deriving Δ. Sheldon (1969) and others (Lieberson 1969; Peet 1974, 1975; Whittaker 1972) identify and address the difficulties surrounding evenness measures. Moreover, given the multiple problems associated with heterogeneity measures and their common use as Δ, the above expressions are considered unacceptable.

Although evenness remains an awkward property to measure, two methods appear to be the least problematic in regard to inherent limitations. The first of these is simply to plot the abundance of species in terms of their rank order from most abundant to least abundant, as suggested by Whittaker (1972). Abundance, as measured by the number of individuals or proportion of total, is plotted on the ordinate axis against the species sequence on the abscissa. The resultant curves and their slope values, commonly called importance values, have been successfully exploited by ecologists (e.g., Lamont *et al.* 1977; Odum *et al.* 1960; Whittaker 1972).

The second method of measuring evenness relates to the moments of the probability density function under study (i.e., mean and standard deviation). Thus, Fager (1972) suggests using the standard deviation of the number of individuals per species in the arithmetic frequency distribution, while Preston (1948, 1962) prefers the estimated standard deviation of the lognormal distribution. Both Whittaker (1972) and May (1981) recommend the variance or its associates as adequate measures of species evenness. In short, a simple estimate of the variance of the proportional abundance of species thus appears well suited for a comparison of the evenness between assemblages.

Heterogeneity

Heterogeneity and its family of measures attempt to simplify the complex relationship between the number of species present and their individual frequencies. The earliest approximation of heterogeneity is illustrated in Simpson's (1949) λ index, as given in equation (15). This expression (15) measures the probability that two individuals drawn at random, with replacement, are representative of the same species (Whittaker 1972). In the following expression, Simpson (1949) defines the probability of interspecific encounter when sampling is without replacement (finite samples):

$$H_1 = \sum_{i=1}^{s} n_i(n_i - 1)/N(N - 1) \quad (21)$$

where n_i is the number of individuals in the ith species and N is the total number of individuals; $N = \sum n$. It is evident that equation (21) is a close approximation of equation (15). Since both λ and H_1 values vary inversely with the heterogeneity of the community, modifications to the equations have been suggested:

$$H_2 = 1 - \sum_{i-1}^{s} \pi_i^2 \quad \text{(Greenberg 1956; Lieberson 1969)} \quad (22)$$

$$H_3 = 1 \Big/ \sum_{i-1}^{s} \pi_i^2 \quad \text{(Williams 1964; Whittaker 1972)} \quad (23)$$

where π_i is the proportion of individuals in the ith species. Pielou (1977) notes that the best statistical approximation for the interspecific probability that two species are different in a finite sample is given by:

$$H_4 = 1 - \sum_{i-1}^{s} \{[n_i(n_i - 1)]/[N(N - 1)]\} \quad (24)$$

A second group of heterogeneity measures have been termed dubious indices by Hurlbert (1971). This group revolves around two fashionable measures that are 'linked by an ectoplasmic thread to information theory' (May 1981:218). As originally defined by Shannon and Weaver (1949), information for infinite populations can be expressed as:

$$H'_5 = -\sum_{i-1}^{s} p_i \log p_i \quad (25)$$

where p_i is the percentage of importance of the ith species. Commonly, H'_5 is estimated for finite populations by:

$$H'_6 = \sum_{i-1}^{s} (n_i/N) \log (n_i/N) \quad (26)$$

where n_i/N attempts to estimate p_i, as the proportion of individuals in the ith species of sample size N. Extensive criticism has been laid against the Shannon and Weaver index; most notably, the problem of sample-size dependence (Pielou 1975, 1977; Smith and Grassle 1977; Smith *et al.* 1979). In support of information theory, Pielou (1975) offers Brillouin's index for finite samples:

$$H'_7 = -\frac{1}{n} \log \frac{\pi_{i-1}^{s} (n_i!)}{(n_i!)} \quad (27)$$

where n_i is the number of individuals in the ith species in a finite sample n. This momentary reprieve for information theory was revoked following Peet's (1974, 1975) results of analysis on the inadequacy of the Brillouin and the Shannon and Weaver indices.

Given the analytical results of Peet (1974, 1975) concerning the inadequacies of the Brillouin and the Shannon and Weaver indices, coupled with the conclusions of Whittaker (1972), the group of fashionable measures appears suspect. Clearly, the use of a single value (i.e., heterogeneity) to describe diversity must be viewed with some caution, since, as emphasized by May (1981), a single value masks the different properties of richness and evenness. Indeed, the inherent difficulties of attempting to contend with two different properties with one value leads to the conclusion that use of the heterogeneity indices should be abandoned. However, with the acknowledgement that heterogeneity measures will continue to be employed, the expression given in equation (24) appears to be the least objectionable.

In our discussion of diversity we have identified and defined three related but distinct components: (1) richness; (2) evenness; and (3) heterogeneity. Richness is viewed as a measure of species variety and three viable methods are suggested to be most appropriate for its determination. The simple species count, also a measure of richness, is known to be biased, given a dependence on sample size (Peet 1974, 1975).

Evenness is viewed as a measure of the proportional abundances of individual taxa. There are two procedures which appear to be the most suitable descriptors of this property. Finally, the third component of diversity, termed heterogeneity, is shown to encompass both richness and evenness under a single value. One widely used example of a heterogeneity measure is the Shannon and Weaver index, also referred to as the information statistic. All heterogeneity measures are considered to be inappropriate at the present time for the purposes of archaeological research. Finally, we contend that the synonymous treatment of the terms 'diversity', 'richness', 'evenness', and 'heterogeneity' is equally unacceptable. In the discussion which follows we review the use of the concept of diversity in archaeological research. This review emphasizes past accomplishments in light of the preceding discussions.

The state of the art

As in most disciplines, archaeologists often profit by exploiting the methods and theories of others. Clearly, the need to interpret human behavior in time and space necessitates a rigorous and occasionally quantitative approach. It is not surprising, then, that attempts have been made toward the integration of diversity concepts into general archaeological thought. Accepting artifactual remains as a paramount source of data, one is in principle receptive to any attempt at recognizing theoretical and structural associations between and within artifact assemblages. In this way, a conceptual substitution of artifact types for species allows researchers to formally apply the concept of diversity to archaeology. In the previous section we briefly introduced and decomposed the concept of diversity as presently understood in ecology. Adopting the principles presented, we shall now review the state of the art. We confine our review to specific uses of the concept of diversity and intentionally ignore those many instances where components of the concept have been used but not identified under the rubric of diversity (e.g., Nance 1981; Tainter 1977b, 1978; Tainter and Cordy 1977).

One of the earliest attempts at integrating diversity into archaeological theory is that provided by Schiffer (1973). Schiffer (1973:114) proposed that '*when the frequency of access to the contents of a facility is either moderate or high, the amount of access volume increases as a function of increasing diversity of contents*' (italics original). To test this proposition Schiffer used the following equation:

$$D_v = \sum_{i=1}^{n} \left(\frac{A_i}{N}\right)^2 \qquad (28)$$

where D_v is the volume diversity, A_i is the number of items in the ith class and N is the total number of items. Since equation (28) is identical to Simpson's expression (15), the former is a measure of species richness. However, Schiffer defined his classes (= species?) as the uniform volume of 100 cm^3 and abundance as number of items per set volume. In other words, he has arbitrarily chosen a sampling unit of 100 cm^3 and is therefore measuring the density per unit volume and not class richness as the equation would imply. By further regressing D_v on percentage access space, Schiffer creates a species area curve. Unfortunately, this results in redundancy since a change in class volume will alter the resultant D_v values but the correlation of D_v to access will remain as a collector's curve.

This example of the use of diversity in archaeology is important as it indicated the degree of sophistication which could be immediately exploited by the archaeological community at large. It is ironic that the earliest archaeological attempt should employ the earliest ecological equation known. Following Schiffer's brief but useful introduction to diversity, a number of years passed before the concept was again applied in archaeology.

Yellen's (1977) classic study on the !Kung owes much of its success to his conclusive results derived from a measurement of artifact diversity. Yellen proposed that

> the longer an area is occupied, the greater the number of activities likely to occur and be repeated there. I guessed that nuclear areas and special activity areas could be distinguished on this basis and sought an index that could quantitatively measure the relative richness of any particular area within a site. I wanted *richness* to be based on two factors: the number of different kinds of remains present and the relative amount of each one. (1977:107)

As Yellen chose to measure both richness and evenness, he correctly opted for a heterogeneity (not richness) measure; in this case the Shannon and Weaver index (equation 25). Unfortunately, his data are defined as finite samples and because the Shannon and Weaver index is designed for infinite populations and is sample-size dependent, a different measure is required. Such finite samples require the use of Brillouin's equation (27), noting all along that all heterogeneity measures are affected by slight changes in richness and evenness so that the resultant values react erratically. A measure of either richness or evenness would be more appropriate. Another difficulty in this analysis is Yellen's tendency to treat bones, grass mats, and stones as taxonomically equivalent classes, and to include arbitrary 10 cm^2 areas as a species equivalent to discrete objects. We note as well that Yellen's periods of site occupation, which range between 5.87 and 6.98 days, may be problematic. Is this time period adequate to obtain a maximum value of community species richness for artifact assemblages? Would a longer period of occupation affect the taphonomy of the artifacts resulting in a balanced evenness value? Although Yellen produced one of the lengthiest examples of diversity application in archaeology, we feel that the final results should be re-evaluated in light of recently developed insights in the mathematics of diversity equations.

Dickens' (1980) study of ceramic assemblages in the South Appalachian ceramic province is unique, given his concern with time-transgressive changes in ceramic diversity. Briefly, he suggests that material traits will peak shortly after a period of 'increased interareal cultural exchange' (Dickens 1980:35). To test his 'Hopwellian interaction' hypothesis, Dickens used the following measure of ceramic diversity:

$$D_w = 1 - S \qquad (29)$$

where $S = \sum_i^n x_i^2$ and x_i is the proportion of individuals (specimens) in the *i*th species (= artifact type). Comparison with our previously defined measures indicates that equation (29) is equivalent to equation (22), which is a heterogeneity measure for the probability of interspecific encounter for infinite populations. Values obtained from this equation increase with increasing diversity, given the inverse effect of subtraction from the constant – one. As noted earlier, if heterogeneity values are thought to be acceptable, the more appropriate equation for Dickens' finite samples is given by equation (24). Nonetheless, we note several positive aspects in his study: (1) emphasis on strict chronologic control; (2) equivalency in species identification; that is, only ceramic artifacts were compared; and (3) his attempt to use samples with more than 500 specimens.

Coeval with Dickens' heterogeneity application on American ceramics, Conkey (1980) published an interesting European example employing Magdalenian bonework. Conkey's primary concern was to show that the site of Altamira differs considerably from other hypothesized dispersion sites in Cantabria. Following Yellen's earlier work, Conkey employed the Shannon and Weaver information statistic to meet the objective of testing an aggregation hypothesis. The underlying argument of her study is unique, that aggregation sites will display greater assemblage diversity than dispersal sites and in this way Conkey's study corollaries add considerably to the study of diversity. In short, she notes that group size, length of occupation, and extent of occupation will have important effects on the resultant observed or measured assemblage diversity. These factors should therefore be a part of all archaeological diversity interpretations.

In his study of archaeological sites near Patoka Lake, Indiana, Cook (1980) relies on Sanders' (1968) research to circumvent problems of sample-size dependency. Cook's attempt was novel; unfortunately the technique used is incorrect. As stated earlier, Sanders' (1968) original rarefaction methodology was revised into correct form in 1971 (Hurlbert 1971) and several times thereafter (Antia 1977; Fager 1972; Heck *et al.* 1975; Raup 1975; Simberloff 1972). Thus, if Cook had applied the proper algorithm (equation (11) above), the example would represent one of the best diversity applications to date.

In a comparison of two assemblages from sites located in the Birch Mountains of northern Alberta, Ives (1981) relies on a number of diversity equations. Ives notes that heterogeneity is composed of two parts: total number and evenness, and then uses the Shannon and Weaver index (equation 26) and the McIntosh index (equation 30) to measure heterogeneity.

$$D_{mc} = \frac{(N - \sqrt{\Sigma n_i^2})}{(N - \sqrt{N})} \qquad (30)$$

Quite correctly, Ives (1981) notes that equation (26) is more sensitive to changes in rare classes, while equation (30) is more sensitive to changes in the abundant classes. As a measure of evenness, Ives proposes use of the Kreb statistic given by:

$$E = \frac{H'}{H_{max}} \qquad (31)$$

which is equivalent to our earlier defined equation (16). Of particular interest in this study is the fact that Ives utilizes different measures to assess different characteristics of the assemblage populations rather than simply select a single measure. We consider this an appealing methodological approach given the variety of population characteristics that may be addressed and the variety of measures available that are suited to these characteristics.

Rice (1981) proposed to test a model of increasing craft specialization through time using ceramic data from the Mayan site of Barton Ramie, Belize. In formulating her test, Rice recognized two components to diversity: richness and evenness. To measure richness she employed the Shannon and Weaver index of heterogeneity given by equation (26) and then measured evenness using Pielou's equation (19). Her application of equation (26) does not require elaboration, as our earlier complaints to similar misapplication apply once again. Note, however, that reliance on equation (19) is also plagued with problems since this measure is subject to a dependence on sample size and richness. Unfortunately, Rice does not provide the resultant data of her richness (read heterogeneity) and evenness computations. Nonetheless, we suggest a visual assessment of her Figures 2 and 3 in relation to the number of sherds examined through the ceramic complexes (her Table 2). The obvious functional dependence (i.e., autocorrelation) of the Shannon and Weaver index and evenness values on sample size negates her final quantitative results and thus limits the validity of her interpretations.

We conclude our review of the first decade of archaeological uses of diversity with Jefferies (1982). Jefferies (1982) examined the relationship between debitage and site location for Woodland sites in northwestern Georgia, to understand the nature of prehistoric activity and adaptation. By assuming 'that the wider the range of activities carried out at a site, the greater the variety of tools required to perform the tasks' Jefferies (1982:114) chose to employ equation (29), and thus equation (22), as a measure of diversity. In certain ways the objectives and underlying premises of this study mimic parts of those provided by Yellen (1977) and Conkey (1980) if earlier diversity application analogues are sought. Still, Jefferies' study is important, given his reliance on debitage as a primary data source, rather than the commonly used ceramic source. In terms of constructive criticisms, use of equations (29) and (22) should be restricted to infinite populations; hence, equation (24) would have been more appropriate. However, this latter equation may also be considered inappropriate, given Jefferies' intention to measure variety; a richness measure is required and not a heterogeneity index. Finally, because the sample sizes used by Jefferies are unequal and range excessively from 136 to 603, the reliability of the resultant values is affected. This latter conclusion is supported by the results of Jones *et al.* (1983), Thomas (1983b) and our own brief review of sample-size dependency that follows.

Future diversity applications

It was not until 1983 that the use of diversity in archaeology was finally examined analytically. As the preceding review has attempted to show, applications of diversity in archaeology during the first decade were research-supportive; that is, archaeologists freely borrowed and applied existing mathematical equations related to diversity to answer specific research questions. Unfortunately, no one considered the concept and its family of measures worthy of being a research problem in archaeology.

The preceding discussions have explored the formal properties of diversity by emphasizing explicit definition of components and the viable measures which accompany these components. Similarly, the review of existing archaeological applications has sought to appraise those applications in light of the formal definitions. In the following discussion we (1) comment on recent archaeological diversity applications which signal future advances; (2) explore the practical limitations of applying the diversity concept; and (3) examine the relationship between artifact typology and species richness.

Given the inherent complexity of the concept of diversity, one usually prefers to emphasize the study of species richness (Whittaker 1972). By studying the number of species, one assumes that the species involved are tangible in the practical sense. If one equates artifact types or site types with the species concept, certain criteria must be met prior to analysis. First, diversity measurement 'requires a clear and unambiguous classification of the subject matter' (Peet 1974:286). In regard to archaeology, this reinforces the idea that researchers must be consistent in their identification of classes and application of typology. Next, the identification of individuals to a specific taxon assumes the individuals are in fact equal. Finally, the recognition of several taxa assumes each taxon is actually different (Peet 1974). For example, a Navajo utility sherd cannot be identified a second time as a Navajo painted sherd. Similarly, there must be an agreement as to what constitutes a kill site, a campsite, or a village. Thus, the initial typology employed by the archaeologist limits the extent of interassemblage analysis that can be carried out. Even more important is the methodology of interassemblage analysis. If diversity assessed for one site relies only on ceramics, this value may not be quantitatively compared to a diversity value from another site which includes lithics and faunal material. Finally, interassemblage diversity comparisons are compounded by simple sample-size constraints.

Using archaeological examples from site assemblages in the Steens Mountain area of southeastern Oregon, Jones *et al.* (1983) correctly demonstrated that simple artifact class richness is dependent upon sample size. Illustrating their argument graphically, these researchers convincingly showed that a high correlation exists in the bivariate relationship between number of tool classes and sample size for 81 sites.

Expanding on the above, we argue that artifactual assemblage interpretations are bounded by the theoretical and practical limits of typology. Figure 2.1 illustrates the *theoretical limits* of all typologies as described by Grayson (1978) and Jones *et al.* (1983). As one increases the number of artifact specimens (N) recovered

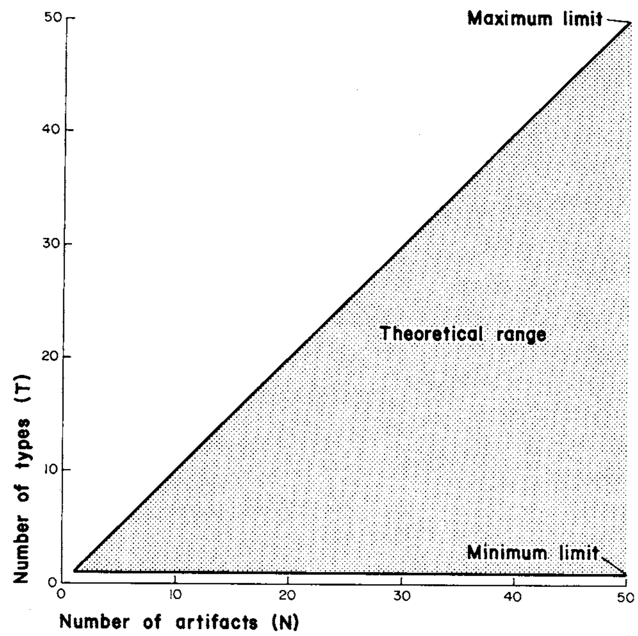

Fig. 2.1. Theoretical range for the number of artifact types (T) as a function of sample size or number of artifact specimens (N)

from a deposit, the concomitant behavior of the types represented by those specimens is restricted. The minimum limit line of Figure 2.1 implies that all specimens recovered will represent a single artifact type. Thus the ratio of types to specimens will be $1/N$. Conversely, the maximum limit line indicates that every new specimen retrieved will represent a new type, in which case the ratio of types to specimens is always one. The actual behavior of a typology within the theoretical range will reflect: (1) the excavation and analytical procedures of individual researchers; and (2) the underlying numerical structure of the assemblage under consideration.

In Figure 2.2 we illustrate the *practical limits* of three archaeological studies. By practical limits we mean the maximum number of types imposed on the assemblage by the researcher, given his or her choice of classification schemes. For example, range A in the figure applies to collections from the lower Chaco River in New Mexico (Reher 1977). Given the raw data in Reher's (1977) study, and the lithic typology adopted in that work, none of the sites examined could exceed a richness value of ten types. Ranges B and C in Figure 2.2 represent the practical limits of the Varangerfjord and Iversfjord (Bølviken *et al.* 1982) regions of Norway, respectively. Bølviken *et al.* (1982) employ differing lithic typologies in the two regions for which maximum richness values of 16 (Varangerfjord) and 35 (Iversfjord) cannot be exceeded by any one site. Individual collections for the three regions must fall within the particular circumscribed practical limits. The actual or observed behavior of collections can now be explored for all three regions.

Using the raw data of number of types and number of specimens for collections itemized in Reher (1977) and Bølviken *et al.* (1982), we generated collector's curves by regressing types on sample size. A natural logarithmic transformation of both variables