

Measuring and Explaining Change in Artifact Variation with Clade-Diversity Diagrams

R. Lee Lyman and Michael J. O'Brien

Department of Anthropology, University of Missouri, Columbia, Missouri 65211

E-mail: lymanr@missouri.edu

Received December 30, 1998; revision received February 19, 1999; accepted March 17, 1999

All archaeologists study and attempt to explain variation in artifacts. Darwinian evolution provides a set of theory-derived methods and explanatory principles that can be used to account for the appearance of novel variants and their differential persistence through time and across space within lineages, or traditions, of artifacts. One method developed by paleobiologists to study the history of the richness of variants within a lineage involves the construction of clade-diversity diagrams. Once an artifact lineage is established, statistical comparison of the artifact diversity represented by that lineage with randomly generated clades provides insights into the origination of novel variants and the history of artifact lineages and suggests explanations of those histories. Such comparisons indicate that (a) the twentieth-century history of diversity of radios attributed by Schiffer (1996) to stimulated variation might be difficult to identify in a prehistoric context; (b) the history of increasing Great Basin projectile point diversity is the result of changes in weapon delivery systems; and (c) the history of increasing diversity in Lower Mississippi Valley ceramic styles is the result of changing patterns of social interaction and transmission. © 2000 Academic Press

INTRODUCTION

Ever since Binford's (1965) criticisms of culture history's "normative theory," archaeologists have focused on numerous kinds of variation evident in the archaeological record. Over the past 2 decades particularly, archaeologists have plotted measures of variation against time (Jones et al. 1983; Kintigh 1984; Leonard and Jones 1989; McCartney and Glass 1990; Meltzer et al. 1992; Rhode 1988; Shott 1989). The tendency has been to present measures of variation as single numerical values, though it was suggested early on that a "histogram of class frequencies ... contains more information than does any single numerical summation of those frequencies" (Dunnell 1989:144). Further, efforts have focused either on determining which of several measures of variation

might be used or on determining if available samples are influencing the measures and how to control those influences. We agree with Dunnell's (1989) observation that minimal effort has been devoted to building theory for explaining fluctuation in variation over time and Schiffer's (1996:655) observation that archaeologists using Darwinian tenets of evolution have been "largely silent about variety-generating processes."

In this article we take initial steps toward filling the voids Dunnell and Schiffer identify. We describe a graphic technique developed by paleobiologists for analyzing and explaining a particular kind of variation as it is expressed over time. Our approach is couched within Darwinian evolutionary theory because we find it to be a fruitful source of analytical models and to have considerable ex-

planatory potential (Lyman and O'Brien 1998; O'Brien and Lyman 2000). Because we have discussed this theory's applicability to archaeological problems in some detail elsewhere [O'Brien et al. (1998) and references therein], we do not repeat those discussions here but take as a given the theory's utility and proceed accordingly.

VARIANTS AND VARIATION

When archaeologists speak of *variation* in the archaeological record, what they typically mean is that there are various kinds of things—that is, there exists more than one variant of a category of stuff, where a variant is a kind differing to some degree from other kinds. Most archaeologists today would agree that the construction of a typology or classification must be problem oriented—the so-called types resulting from the construction must be related to the research questions being asked. For example, use-wear on stone tools tells us about the function of those tools; thus, attributes of use-wear rather than, say, the color of the objects are used to build functional types. Other kinds of attributes are used when archaeologists desire to construct historical types of stone tools to measure the passage of time (e.g., O'Connell and Inoway 1994; Thomas 1981).

Once artifacts are classified, variation is then measured. Potential measures, or indices, of variation are numerous, though archaeologists have tended to focus on three that capture in one way or another the structure of that variation. Measures of richness are simply tallies of the number of kinds or variants in a set of phenomena. Measures of evenness summarize the frequency distribution of specimens within the multiple kinds of a classification or typology. A collection in which the constituent members are distributed equally across all kinds is more

"even" than a collection in which the members are abundant in one kind, of mid-level abundance in another kind, and rare in a third. Measures of heterogeneity represent a combination of richness and evenness and can be calculated in various ways [see chapters and references in Leonard and Jones (1989)].

Regardless of which measure is used, the *value* taken by that measure of variation is a function of the classification system used. Thus as Dunnell (1989:145) notes, "the meaning of diversity indices is mechanically determined by the definitions of the categorical inputs [read *classification units*]; if the [classification units] on which diversity indices are based do not unambiguously measure the parameter of interest, then no interpretations are warranted." If one wishes to know how many species of mammal are represented in modern Iowa, then one must know what a species is, what a mammal is, and what an organism is so that one considers only organisms that are mammals and the categories tallied are species. As Jones (1989: 304) points out, not only must our classification be "constructed with a purpose and have meaning warranted by archaeological theory," the classes or types "must be mutually exclusive, exhaustive, and on the same [taxonomic] level, ensuring that classes are equally different in a structural sense."

The importance of how variation among artifacts is sorted and classified cannot be overemphasized. Usually, variants or types are specified for an analytical purpose. Measures of richness (which we use here as synonymous with diversity), evenness, and heterogeneity of collections are meaningless if the types underlying the measures were designed for disparate purposes. In the next section we discuss two kinds of variation that might be measured, one of which is critical to understanding the history of variation and to the analytical technique we describe.

THE GENERATION AND TRANSMISSION OF VARIATION

Archaeologists have traditionally called on ethnographically documented mechanisms of variation generation such as invention, innovation, and diffusion. In part because ethnographic informants answer questions regarding why they produce novel variants, archaeologists have also spoken of the "intentional," "guided," and "directed" production of novel variation (e.g., Braun 1991; Neff 1992; Rosenberg 1994; Spencer 1997). These terms imply that if a novel variant was desired, perhaps because it would solve an adaptive problem, it was invented. Although we endorse the notion that new variants are intentionally created at least some of the time (e.g., Lyman and O'Brien 1998; O'Brien and Holland 1992), we have yet to determine how such intentions are to be identified analytically in the archaeological record. Further, as we and others have argued (Dunnell 1989; Leonard and Jones 1987; Lyman and O'Brien 1998), using human intentions as an *explanation* places cause in the phenomena requiring explanation rather than in theory. As a result, it denies an opportunity to understand the evolutionary origins of human intentions. What mechanisms, then, might be called on to account for the generation of novel variants?

Stimulated Variation

Schiffer (1996:655) hypothesizes that "some bursts of variety generation" can be accounted for as follows: "[I]nformation (as matter or energy) coming from changed conditions in selective contexts can stimulate an increase in inventive activities of behavioral components and can foster the creation of new behavioral components." Labeling this "process" "stimulated variation," Schiffer (1996:655) notes, "in no way obviates [natural] selection;

after all, every variant produced during an instance of stimulated variation can be selected against." Further, it "allows us to assign problem-solving its proper role in evolution, that of producing new variants" (Schiffer 1996:655). To avoid the connotation of "directed" or "guided" variation, Schiffer (1996:655) states that a novel variant "created by [stimulated variation] is not directed by future adaptive needs, but is shaped by contemporaneous phenomena in the selective environment."

In our view, the key word is "stimulated" because it implies two things. First, it implies that at least some novel variants will appear as a result of new *perceived* needs, that is, as a result of intention (e.g., Basalla 1988; Petrosky 1992; Stein and Lipton 1989). Hence we elsewhere refer to "stimulated variation" as another expression of "necessity is the mother of invention" (O'Brien et al. 1998:493). Apparently, regardless of whether all or some of the new variants are selected against, "problem-solving behaviors *usually* lead to an appropriate response, and the result is a new adaptation" (Schiffer 1996:655; emphasis added). At least some new adaptations will, therefore, be the result of problem-solving activities that were stimulated by perceived changes in selective contexts. Second, stimulated variation implies that the generation of variation must be restricted to a cultural lineage—change must be internal to an evolutionary line—and that the source of the novelties may be internal (invention, innovation) or external (diffusion) to the lineage.

Identifying instances of stimulated variation has steep data requirements. First, Schiffer (1996:656) notes that invention is a "major source of variation [and] is hardly a random process; rather, inventive activities can be highly patterned by stimulated variation. These strong effects are often discernible as a clustering in time and space of similar inventions." This characterization indicates that a first step

in the identification of instances of stimulated variation requires that the random background generation of novel variants must be distinguished from the spatio-temporally restricted instances of the generation of multiple novel variants—what Schiffer (1996:656) describes as “dramatic increases in variation” or a “burst of variation”—signifying stimulated variation. We describe below both a mechanism that we believe randomly produces novel variants and an analytical technique to distinguish between that kind of variation generation and stimulated variation.

Schiffer (1996:656) suggests that the new variants resulting from stimulated variation will be “rigorously winnow[ed]” by selection—they will be “shaped by the selective environment” (Schiffer 1996:655)—and thus only some may persist over time. This underscores that classifications must be theoretically informed and must monitor variation in great detail, but also that our classifications must monitor variation in two rather different ways. First, classifications must monitor variants that influence adaptive fitness and thus are subject to—“shaped by”—natural selection. We term these “functional variants” (Dunnell 1978). To identify, monitor, and explain this kind of variant we need detailed knowledge of selective environments and their histories. Then, fine-scale spatiotemporal correlation of changes in selective environments and “dramatic increases” (Schiffer 1996:656) in the diversity of novel functional variants must be found. Given sufficient temporal resolution, the correlation will be less than perfect in a predictable way because, according to the definition of stimulated variation, the stimulus must temporally precede the appearance of novel functional variants.

A Darwinian perspective acknowledges that novel variants may not influence adaptive fitness but rather be adaptively neutral. This is the second kind of variation that must be identified and monitored; we label these “stylistic variants” or

simply “styles” (Dunnell 1978). Styles are adaptively neutral and thus are not subject to natural selection—they are not “shaped by the selective environment” (Schiffer 1996:655). Stimulated variation can result, as Schiffer (1996:656) notes, in “false starts, partial solutions, unintended consequences, and dead-ends.” All these potentialities, as well as the persistence of variants that have no apparent adaptive advantage, may be explicable as selectively neutral variants of various scales rather than as variants whose history subsequent to their initial appearance is “shaped by the selective environment.” Thus, style as a particular kind of variation (as we define it) has analytical utility.

There is also the necessity of recognizing styles. Stimulated variation occurs within a cultural lineage, and one must, therefore, somehow establish that a lineage, or line of heritable continuity (signifying transmission), is under study rather than merely a sequence of variant forms. As Neff (1993:27) indicates, “Detecting transmission is a matter of ascertaining temporal and spatial boundaries and amassing data on the distribution of [artifact] traits arising from shared [artifact]-making practices. . . . Similarities may arise through convergence in the absence of cultural transmission, and these analogous similarities may be confused with evidence of cultural relations.” In our view, stylistic similarity of artifacts is “homologous similarity [that is] the result of direct cultural transmission once chance similarity in a context of limited possibilities is excluded” (Dunnell 1978:199). We return to this below and describe a technique for testing similarities that are hypothesized to comprise homologs.

Transmission Fidelity and the Random Generation of Background Variation

The concept of stimulated variation describes an important mechanism for the

generation of novel variants. It appears to be incompletely characterized by Schiffer (1996), however, as it tends to focus on what we term *functional* variants. It does not account for the persistence of adaptively neutral variants, and it excludes what we term the background, or random generation, of novel variants. All these factors are included within modern Darwinian theory, particularly that portion dealing with transmission and heritability. In short, the *fidelity* with which a thing is replicated varies from a perfect copy to a copy that is to a greater or lesser degree an imperfect copy of the original. Without using the term "fidelity," Kroeber (1940), writing on what he termed "stimulus diffusion," or "idea diffusion," noted that cultural transmission could involve perfect or imperfect fidelity. Forty years later, zoologist G. E. Hutchinson (1981) noted that various organisms, including humans, learn by mimicking and imitating the behaviors of those conspecifics around them. Importantly, Hutchinson (1981:164) stated that the imitator is a potential innovator: "In all kinds of imitation, errors of communication are very likely to occur. They are to mimesis what mutation is to ordinary [genetic] inheritance." We find transmission and its errors a significant potential explanation for the origination of novel variants (Lyman and O'Brien 1998; see also Neff, 1992, 1993) and suspect that it accounts for much of the random background generation of new variants. Errors of transmission—incomplete or incorrect message sent or received—result in new empirically expressed variants rather than in replication of the original variant. Within a lineage, new variants produced by transmission errors should be more or less randomly distributed over time and space.

Transmission also accounts for the persistence of selectively neutral variants or styles. Histories of stylistic variants illustrate the vagaries of transmission, and

computer-generated simulations of transmission by archaeologists (e.g., Lipo et al. 1997; Neiman 1995) and paleobiologists (e.g., Gould et al. 1977; Kitchell and MacLeod 1988; Raup et al. 1973; Raup and Gould 1974; Uhen 1996) show that selectively neutral variants will persist differentially and that we shouldn't expect patterned or significant increases or decreases in stylistic diversity over time. The frequency distributions of stylistic variants over time are "stochastically propelled" (Dunnell 1978:199; O'Brien and Holland 1990:53). Because styles are selectively neutral, fluctuations in the frequency of stylistic variants have a Markovian structure simply because of "inheritance" (Dunnell 1980:58) or transmission—that is, what comes later depends in part on what came earlier (Gilinsky and Bambach 1986). Whether the idea for a cultural variant—of whatever scale—is transmitted is stochastic because it depends on history. The variant must exist (even if only conceptually) to be transmitted, but it does not have to be transmitted merely because it exists. And if it is transmitted, it does not have to be replicated or be empirically expressed. This applies to functional variants as well as to stylistic variants.

Transmission is critical to the evolution of a lineage because it can result in the differential persistence of variants, whether those variants are stylistic or functional. This is clear given the concept of *sorting*. Vrba and Gould (1986:217) indicate that sorting comprises "differential birth and death among varying organisms [variants] within a population" and that it "is a simple description of differential representation; it contains, in itself, no statement about causes. At its core, Darwinism provides a theory for the causes of sorting—natural selection." Their point is that change over time either in the frequency of variants or in the frequency of specimens within particular variants may be

the result of selection, but it may also be the result of nonselection-driven sorting (Gould 1988). For example, the latter can occur when selection operates at one scale—say, on a kind of discrete object—and results in the differential persistence of phenomena at a more inclusive or less inclusive scale—an aggregate of particular variants or an attribute of a variant, respectively. In other words, selection works directly on one scale of phenomena but only indirectly on another scale; the former is selected, whereas the latter is merely sorted. Other sorting mechanisms include drift (sampling error that occurs during the transmission process), random extinction events, and the hitchhiking, or mechanical association, of a trait with another that is being shaped by selection [Gould 1990:21; see Vrba (1989) for extended discussion]. Identifying instances of sorting and distinguishing them from instances of selection requires not only the same kinds of data demanded for identifying instances of stimulated variation but also a detailed examination of phenomena at various scales.

ESTABLISHING CULTURAL LINEAGES

The transmission process underlies efforts to establish cultural lineages, or lines of heritable continuity, and archaeologists have at their disposal an analytical technique for constructing and testing whether a temporal sequence of artifact forms comprises a lineage. That technique is frequency seriation, which has seen detailed discussion at the hands of evolutionary archaeologists in the past several years (e.g., Lipo et al. 1997; Neiman 1995; Teltser 1995). Thus we only summarize its history and most critical aspects here [see Lyman et al. (1997, 1998) and O'Brien and Lyman (1999, 2000) for additional details].

Evolutionary Affinity among Artifacts

Wiley (1953:363) stated as an unequivocal methodological axiom of archaeology that "typological similarity is an indication of cultural relatedness." Evans (1850) used this axiom in his phyletic seriations (Lyman et al. 1997, 1998; O'Brien and Lyman 1999) of British gold coins, Petrie (e.g., 1899) used it to accomplish his phyletic seriations of Egyptian pottery, and Kidder (1917) and Beals et al. (1945) used it in their phyletic seriations of Southwestern pottery. The axiom also allowed archaeologists to perform "typological cross dating" (e.g., Patterson 1963), and it prompted development of the concept of "horizon styles" (Kroeber 1944; Willey 1945), both of which assume that typologically similar artifact *styles* are close in time. The emphasis on style is critical. Although there was no explicit theoretical explanation of stylistic similarity, the seldom-remarked notion was that formal stylistic similarity denoted a close phylogenetic relationship in a sense analogous to that in biological evolution (Lyman 1998)—artifacts are stylistically similar as a result of *cultural* transmission.

Archaeologists in the first half of the twentieth century argued about how to determine whether *typological similarity* denoted common evolutionary history—similarity was homologous—or functional convergence—similarity was analogous—but with little in the way of agreement (Lyman 1998). In fact, it was this debate that Binford (1968:8) used to illustrate the weakness of the culture-history paradigm: "It is evident that each culture trait tabulated in obtaining the ratio which measures degree of genealogical affinity must be evaluated to determine whether the similarity between traits arose as a function of lineal transmission, diffusion between cultural units, or independent development within each cultural unit. It is here that a basic, unsolved problem lies:

How can archeologists distinguish between homologous and analogous cultural similarities?"

The frequency-seriation technique comprises a way to test hypothesized homologous similarities (Dunnell 1970; Lipo et al. 1997; Neiman 1995; Teltser 1995). Types that are homologous will display particular frequency distributions over time and space; types that do not display such distributions are not homologs. The analytical focus of frequency seriation is on changes in relative frequencies of specimens representing each of multiple stylistic variants. One can build and test a hypothesized cultural lineage with the frequency-seriation technique and then monitor change in diversity of either stylistic variants or, after reclassifying the artifacts in functional terms, functional variants. The requisite first step, however, is in building and testing the lineage, and that requires standard frequency seriation of artifact styles to ensure that heritable continuity is being monitored rather than a mere sequence of variation that may or may not entail heritable continuity.

Seriation and Heritable Continuity

Frequency seriation attempts to deal strictly with heritable, or phylogenetic, continuity through three procedural requirements (Lipo et al. 1997; Teltser 1995). Although the roots of the requirements are deep (e.g., Ford 1938; Rouse 1939), one of the earliest and most detailed statements on them is found in Phillips et al. (1951:219-236). The requirements were expanded, amended, and clarified in later years (e.g., Cowgill 1972; Dunnell 1970; Ford 1962; Rouse 1967; Rowe 1961). First, and perhaps least important, given the focal concern of phylogenetic continuity, the assemblages of artifacts to be seriated must be of similar duration. Meeting this requirement insures that the placement of particular assemblages in an ordering is the result of their

age and not their duration. The other two requirements are critically important to the issue of phylogenetic relatedness. The second requirement is that all assemblages to be ordered must come from the same local area. Meeting this requirement, though an analytically complex procedure (Lipo et al. 1997; see also Dunnell 1981), attends the fact that heritable continuity has both a temporal and a spatial component (Teltser 1995) and attempts to control for the latter in an effort to measure only the former.

In our view, meeting the second requirement increases the probability of meeting the third requirement [following, particularly, Rouse's (1955) reasoning], which is that the assemblages to be ordered in a frequency seriation must all belong to the same cultural tradition. A cultural tradition is defined as "a (primarily) temporal continuity represented by persistent configurations in single technologies or other systems of related forms" (Willey and Phillips 1958:37) or as "a socially transmitted form unit (or a series of systematically related form units) which persists over time" (Thompson 1956:38). Therefore, if one meets the third requirement of the seriation method, then heritable continuity is assured and phylogenetic affinities between the seriated assemblages are guaranteed. The third requirement means that the seriated assemblages "must be 'genetically' related" (Dunnell 1970:311; see also Ford 1938; Kidder 1916). The use of theoretical units, or classes, such as are demanded by frequency seriation, satisfies this requirement (Dunnell 1970).

Frequency seriation monitors transmission and heritability at two levels (Rouse 1939). First, each artifact identified as a member of a particular class is *hypothetically* related phyletically to every other specimen within that class, given their properties in common [they are members of the same class (type)] and, typically (though not necessarily), their spatiotem-

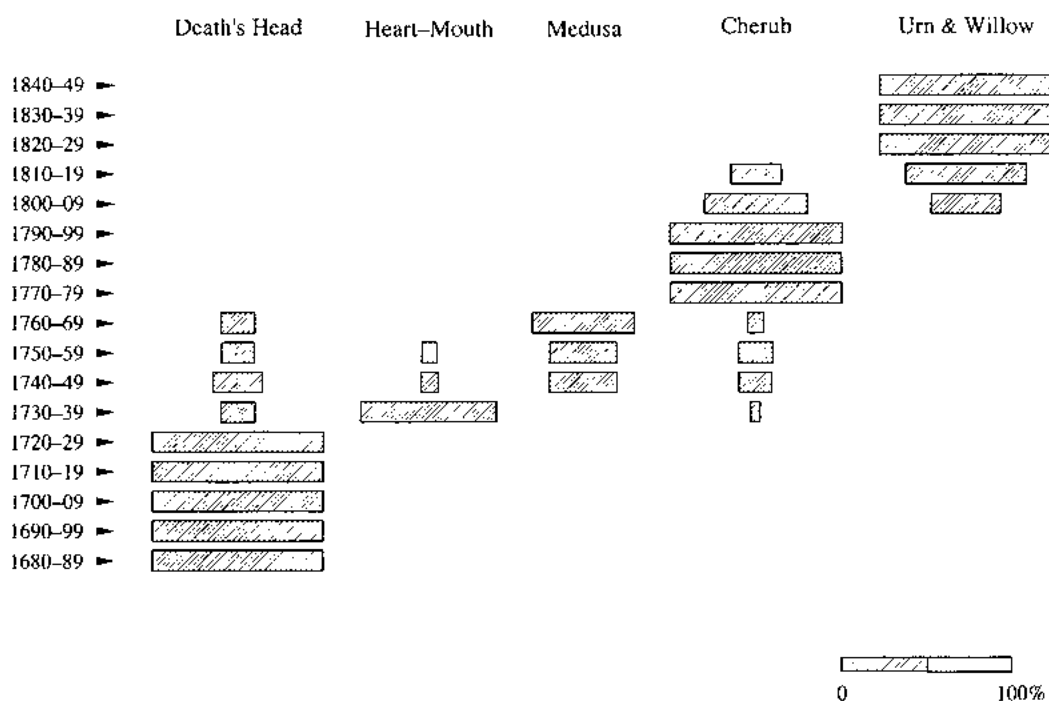


FIG. 1. Edwin Dethlefsen and James Deetz's graph showing changing percentages of five classes of headstones in use in Plymouth, Massachusetts between 1680 and 1849 (after Dethlefsen and Deetz 1966:Fig. 3c).

poral propinquity (e.g., Phillips et al. 1951; Rouse 1955; Steward 1929). We hereafter refer to this as the *type/species* sense of heritable continuity. Second, the multiple classes that are seriated are *hypothetically* related phylogenetically, given the requirement of seriation (Dunnell 1970; Lipo et al. 1997) that all seriated collections derive from a single cultural tradition, which, by definition, reflects transmission, persistence, and heritable continuity (Phillips and Willey 1953; Thompson 1956; Willey 1945; Willey and Phillips 1958). Because traditions can be conceived of and constructed at the scale of attribute of a discrete object, type of discrete object (a particular combination of attributes), or multiple types of discrete objects (e.g., Neff 1992), we hereafter refer to this as the *tradition/lineage* sense of heritable continuity to signify the potential for a diversity of

units—of whatever scale—within a tradition or lineage.

We emphasize that heritable continuity at both the *type/species* level and at the *tradition/lineage* level is *hypothetical*. This means that the phylogenetic relationships of the seriated materials are testable. If the requirements of seriation are met, then the frequency distribution of each class over time will, as a result of transmission (Lipo et al. 1997; Neiman 1995), display a unimodal curve such as that shown in Fig. 1. That is the implication of frequency seriation as a test of heritable continuity. The use of classes of artifacts representing what are referred to as "historical," or "temporal," types (styles) insures heritable continuity at the *type/species* level—items are definitionally identical—and, with appropriate specification of the set of classes to be used, at the *tradition/lineage*

level as well. Two points need to be made in this respect. First, the distribution of styles should, theoretically, be different from the distribution of functional forms over time and space. The former will measure interaction, transmission, and inheritance, whereas the latter will sometimes measure transmission as mediated by natural selection and other times measure adaptational change alone (Meltzer 1981). In other words, functional forms *might* display discontinuous, or multimodal, frequency distributions over time as a result of, say, convergence or fluctuations in selective environments; when they display such distributions, they cannot serve as historical types.

The second point with respect to stylistic variants is that the phylogenetic implications of the hierarchical structure of the Linnaean taxonomy in biology are transferable to a similar hierarchical alignment of historical types of artifacts. Although a hierarchically structured typology need not have phylogenetic implications (Valentine and May 1996), a significant part of evolution comprises diversification, and thus a hierarchy of successively more inclusive types could ultimately reflect phylogeny, though this too requires testing (e.g., Lipo et al. 1997). The first pots manufactured were no doubt less diverse in kind than were their technological descendants. Further, differences between lithic technology and ceramic technology suggest that they evolved independently of one another, and each therefore can serve as a test of whether the ordering produced by the other reflects the passage of time (Dunnell 1970). A confirming test would also suggest that the orderings comprise lineages. Plotting relative frequencies of types against their stratigraphic provenience comprises what we have termed "percentage stratigraphy" (Lyman et al. 1997:52; Lyman et al. 1998; O'Brien and Lyman 1998, 1999). *If* the types display unimodal frequency distri-

butions and the requirements of the seriation method are met, percentage stratigraphy would not refute the inference that the ordering is chronological and would suggest the sequence comprises a lineage.

Although he doubted that artifact types could ever be used to build a phylogeny, Krieger (1944) noted that temporal types must (a) "serve as tools for the retracing of cultural developments and interaction" (p. 272); (b) have "*demonstrable historical meaning*" (p. 272); (c) comprise "*specific groupings of structural features which have proved historical significance*" (p. 273); and (d) "*occupy . . . definable historical positions[s], that is, [their] distribution[s] are] delimited in space, time, and association with other cultural material*" (pp. 277–278). That such units had to be constructed by "trial and error" was implicit in Krieger's discussion and explicit in the discussion of others (e.g., Ford 1962; Phillips et al. 1951). This process of constructing historical types goes on today (e.g., O'Connell and Inoway 1994; Thomas 1981), but until recently there has been little concern for *why* these types allow the tracking of Krieger's "cultural developments and interaction" and why they have spatiotemporally limited distributions. Only with explicit adoption of the tenets of Darwinian evolutionary theory has it become clear why historical types behave the way that they do—that is, why they have a particular kind of spatiotemporal distribution (e.g., Lipo et al. 1997; Neiman 1995; O'Brien and Lyman 1999, 2000; Teltser 1995). When they find empirical expression, that behavior is the result of the transmission of selectively neutral variants, or styles.

CLADE-DIVERSITY DIAGRAMS

For nearly a century Americanist archaeologists have graphed change over time in the relative frequencies of specimens within types (Lyman et al. 1998). The

graphs of centered and stacked bars pioneered by Rouse (1939) and perfected and popularized by Ford (e.g., 1949) represent the most familiar example (Lyman et al. 1998; O'Brien and Lyman 1998). These graphs have been and still regularly are used to measure the passage of time. They measure time best when they monitor transmission and heritable continuity (O'Brien and Lyman 1999), but they also monitor the history of variation in artifacts. One could, for example, calculate a richness value, an evenness value, a measure of heterogeneity, or some combination of these for each row of bars in a graph. In the following, we examine only the richness property, but in a very specific way. In particular, we take advantage of the fact that paleobiologists have, since early in the twentieth century, produced graphs that appear similar to those of archaeologists [Gould et al. (1987) and references therein]. It is the graphic technique of paleobiologists that we describe below.

Graphing Diversity among Phylogenetically Related Taxa

Paleontologists, under the term "biostratigraphy," have long used particular fossil taxa to correlate strata based on the notion that "the closer the relationship of two species, the closer they will approximate each other in time" (Eldredge and Gould 1977:39), with closeness of taxonomic relationship being denoted by morphological similarity. Originally, formal similarity denoted only close relationship in time and allowed the correlation of horizontally separate strata based on their biological content, similar to the typological cross-dating of archaeology. The evolutionary implication that formal similarity implies a phylogenetic relationship was explicit only after Darwin (1859:206) argued that "By unity of type is meant that fundamental agreement in structure,

which we see in organic beings of the same class, and which is quite independent of their habits of life. On my theory, unity of type is explained by unity of descent." It is now axiomatic that caution must be exercised when determining phylogenetic affinity to ensure that formal similarity is not the result of evolutionary convergence [e.g., Lyman (1998) and references therein].

Paleobiologists construct graphs showing the frequency of lower-level taxa within a particular higher level taxon in order to display the history of life. Many of these graphs produced in the last several decades comprise what are termed clade-diversity diagrams and are constructed as shown in Fig. 2 (Gould et al. 1977; Raup and Gould 1974; Raup et al. 1973). A clade-diversity diagram displays the fluctuating absolute frequency of taxonomic richness within a monophyletic lineage, or clade, over time. What is graphed is the number of taxa within a higher taxon—the number of orders within a class, the number of families within an order, and so on. A monophyletic lineage comprises one or more lineages or taxa having a common ancestor; new lineages or taxa arise only by branching or diversification, and each clade contains only and all those taxa deriving from a common ancestor plus the common ancestor. Because a clade-diversity diagram shows change in the number of extant lineages over time, it comprises a history of the origination and extinction of taxa within a clade: "If the rate of branching exceeds the rate of extinction . . . the number of coexisting lineages increases. If extinction exceeds the rate of branching, diversity decreases. If the two rates are equal over a period of time, diversity remains constant" (Raup et al. 1973:528). Often, hypothesized phylogenetic connections between clades are shown (Fig. 3). In such cases, each column of centered and stacked bars depicts, say, the number of

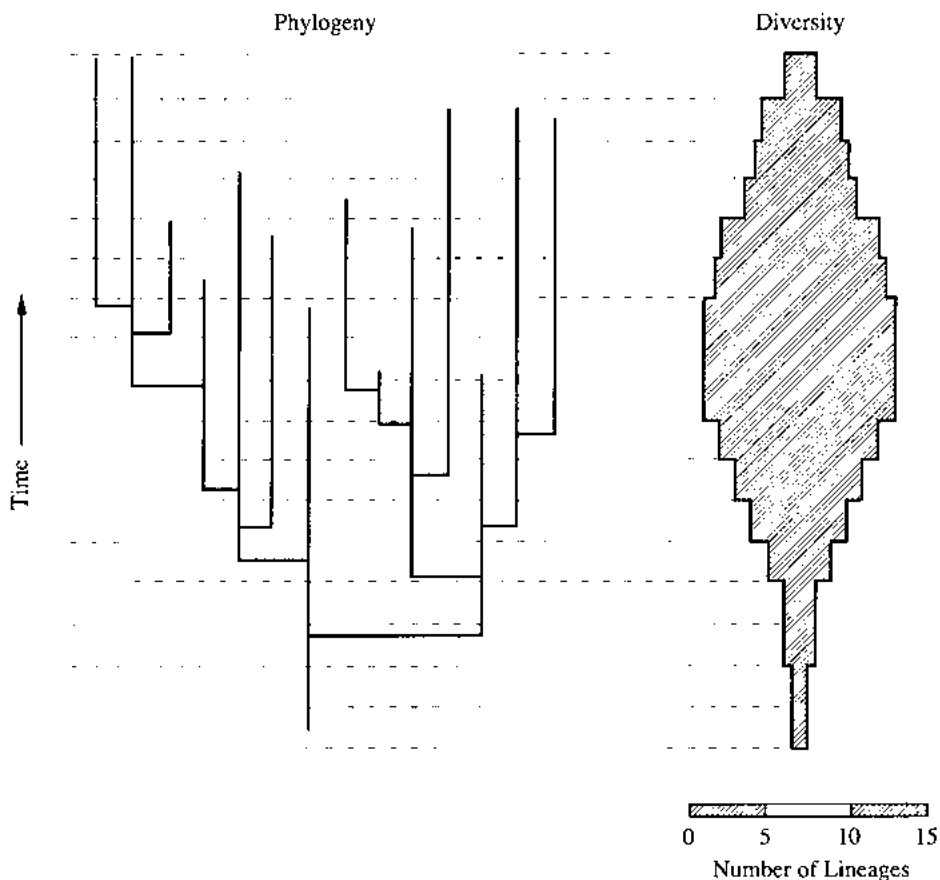


FIG. 2. A model for producing a clade-diversity diagram. The phylogenetic history of taxa is shown on the left and the resulting clade-diversity diagram is shown on the right. Once the phylogenetic history of related taxa has been worked out, including the placement of their starting and ending points, the data are summarized as at the right, which shows the waxing and waning of the number of taxonomically distinct lineages (taxa at a lower, less inclusive hierarchical level) within the clade (after Raup et al. 1973:Fig. 1).

genera within a family, and the phylogenetic relations of the families can be shown with dotted lines because all graphed families belong to the same order or class.

Clade-diversity diagrams and frequency-seriation graphs are similar in three ways: Time passes from the bottom to the top, having typically been determined for both kinds of analysis based on stratigraphy or radiometric dating; frequencies of kinds of things are denoted by vertically stacked, horizontally centered

bars of various widths; and the things graphed are thought to be phylogenetically related. But the frequencies graphed are absolute numbers of kinds of things in a clade-diversity diagram; in a frequency-seriation graph the relative frequencies of specimens within each of several types of a kind of artifact are represented. A clade-diversity diagram can be easily generated from a frequency-seriation graph. One has but to tally up how many types occur in each temporal interval and then generate the appropriate graph of centered and

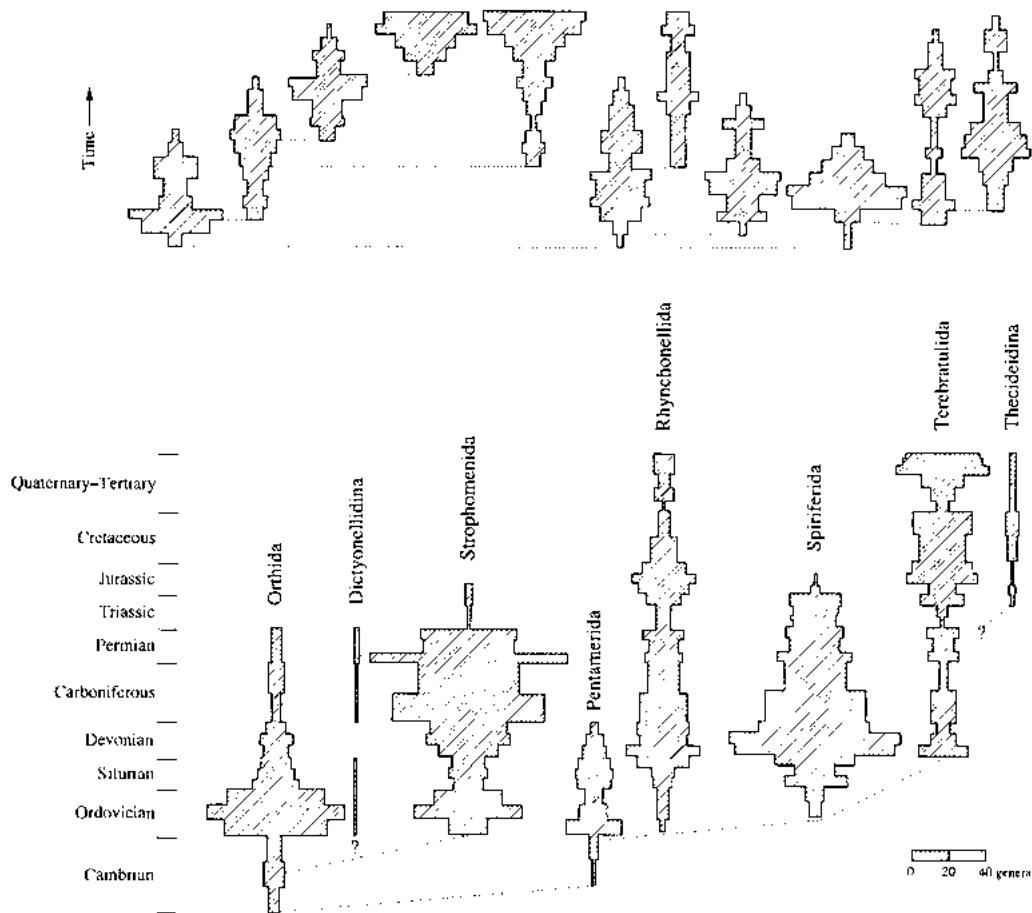


FIG. 3. Clade-diversity diagrams of randomly generated clades (top) and real clades showing the diversity of genera within orders of brachiopods (bottom). The Markovian structure of the graphs depends on heritable continuity (after Gould et al. 1977:Fig. 1).

stacked bars. For example, the clade-diversity diagram in Fig. 4a is based on the graph in Fig. 1 and on the assumption that the latter comprises a single clade.

Clade-diversity diagrams have been compiled for a number of taxa and time periods and have revealed not only various evolutionary events that took place during the history of life (e.g., Prothero 1985; Sepkoski et al. 1981) but the potential causes of those events (e.g., Miller and Sepkoski 1988). Because both simulated and real clade-diversity diagrams have Markovian structures, variations in the computer programs used to generate sim-

ulated clade-diversity diagrams have yielded insights into possible evolutionary processes (e.g., Benton 1995; Gilinsky and Bambach 1986; Uhen 1996).^{1*} Two points are important. First, the same kinds of questions, analyses, and answers can be part of archaeology precisely because it, too, often is faced with apparent evolutionary trends, though in artifactual rather than in organismic variation. Second, analysis of clade-diversity diagrams allows the distinction to be made between

* See Notes section at end of article for all footnotes.

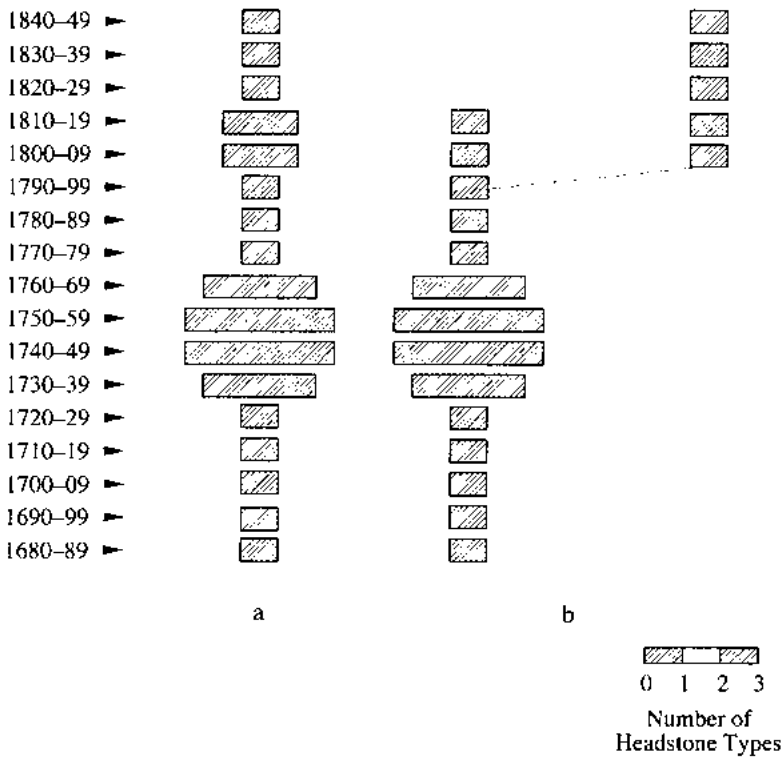


FIG. 4. Clade-diversity diagrams for headstones, based on Fig. 1. (a) Number of classes per time period assuming one clade or tradition; (b) number of classes per time period assuming two clades or traditions (slate-gravestone at left and urn-and-willow at right) and phylogenetic diversification of one from the other (see text for discussion).

stimulated variation and randomly generated background variation.

Distinguishing between Stimulated Variation and Random Background Variation

Paleobiologists developed computer programs to simulate the diversity history of a clade because they "wished to predict what phylogeny would look like if it were determined by random processes and then to compare this with the real world, to enable us to separate random elements from those that require interpretation in terms of specific and perhaps nonrecurring causes" (Raup et al. 1973:527). Using more explicit wording, Raup (1977:53) indicates that "the first question for the pa-

leobiologist faced with an evolutionary trend should be: Does the trend represent a statistically significant departure from chance expectations?" This question must be asked, as Gould et al. (1977:24) note, if explanations are to result: "Until we know what degree of apparent order can arise within random systems, we have no basis for asserting that any pattern in the history of life implies a conventional [typically functional or adaptive] cause for its generation."

Gilinsky and Bambach (1986:251-252) summarize possible conclusions if the question were answered negatively: "Beyond the orderliness necessarily imposed by the system, major features of the history of diversity are truly (ontologically)

random; or . . . major features of the history of diversity result from large numbers of complexly intertwined causal factors none of which predominates, such that [clade histories] often appear to be random." The "orderliness necessarily imposed by the system" can be summarized as follows:

Initial diversity is always zero (before the taxon's inception); final diversity is always zero (after extinction); diversity never swings below zero; and diversity is Markovian (diversity at any time $t + dt$ depends in part upon the diversity at previous time t). These features limit the range of possible diversity histories and cause patterns of diversity change through time to appear more orderly than in independent-events processes, which have no Markovian memory and have no constraints at their beginnings and ends (Gilinsky and Bambach 1986:251).

Gould et al. (1987:1437) posed the question, "[D]oes any asymmetry exist, statistically defined over large numbers of lineages, in the vertical [temporal] dimension of clade diversity diagrams?" If so, one might measure time with the diagrams or detect a recurrent pattern in the history of changes in life's diversity. They found such an asymmetrical pattern in the form of "bottom-heavy clades" and suggested that there indeed was a particular direction to biological evolutionary time [but see Uhen (1996)]. Many of the real clades they examined appeared to be taxonomically richest early in their history and to become progressively less rich over time [but see Kitchell and MacLeod (1988, 1989)]. Importantly, their methods of analyzing clade-diversity diagrams included a technique that can be used to identify Schiffer's (1996) "bursts of variation" characteristic of stimulated variation and to distinguish them from the random background generation of variation.

Gould et al. (1977) developed a simple way to determine if a clade-diversity diagram was symmetrical or not, and if not, whether the diagram was bottom or top heavy. To do this, they determined the "center of gravity," or what they referred

to as the "relative position in time of the mean diversity" of a clade (Gould et al. 1977:26, 1987:1438). The duration of a clade is measured on a scale from zero—the time period immediately prior to the period when the clade first appears—to one—the time period immediately following the period of clade extinction. An equilateral diamond-shaped clade-diversity diagram is symmetrical and has a center of gravity of .5; a tear-drop-shaped, or bottom-heavy, diagram is asymmetrical and has a center of gravity of less than .5; an inverted tear-drop-shaped, or top-heavy, diagram has a center of gravity of greater than .5. The center-of-gravity value is calculated with the formula.

$$CG = (\sum N_i t_i) / (\sum N_i),$$

where N is richness per time interval and t is the scaled temporal position of the richness measure. For example, the center of gravity of the clade-diversity diagram in Fig. 4a is calculated as shown in Table 1. We assume for purposes of illustration that the clade began in 1680 and ended in 1850, but note that this is procedurally incorrect because the temporal end of the clade is unknown. One should not calculate a CG value for an incomplete clade-diversity diagram because if either end is missing, the CG value will be skewed. A top-heavy value will result from a missing terminal portion; a bottom-heavy value will result from a missing initial portion (Gould et al. 1977, 1987; see also Gould 1978). We use the graph in Fig. 4a here merely to illustrate the technique of calculating CG and return to the issue of whether it in fact represents a clade below.

Kitchell and MacLeod (1988, 1989) note that the categories "symmetrical," "bottom heavy," and "top heavy" are nominal scale. Only a CG of .5 denotes the first category, CG values less than .5 denote

TABLE 1

Mathematics Used to Calculate the Center of Gravity Statistic for the Clade-Diversity Diagram shown in Fig. 4a

Time period	Richness (N.)	Scaled time (T.)	Richness × scaled time
1850	0	1.000	0
1840-1849	1	.944	.944
1830-1839	1	.889	.889
1820-1829	1	.833	.833
1810-1819	2	.778	1.556
1800-1809	2	.722	1.444
1790-1799	1	.667	.667
1780-1789	1	.611	.611
1770-1779	1	.556	.556
1760-1769	3	.500	1.500
1750-1759	4	.444	1.776
1740-1749	4	.389	1.556
1730-1739	3	.333	.999
1720-1729	1	.278	.278
1710-1719	1	.222	.222
1700-1709	1	.167	.167
1690-1699	1	.111	.111
1680-1689	1	.056	.056
1679	0	0	0
Σ =	29	—	14.165
CG	14.165 : 29 = .488		

disappear during a time interval) influence the confidence intervals bracketing the mean CG. Thus, there is much that could be done with simulating the history of artifact diversity. Rather than pursue that here, we compare CG values observed among several sets of artifacts with Kitchell and MacLeod's (1988) simulated-sample mean CG of .5 and standard deviation of .032 ($n = 1000$); all p values are for one-tailed tests. Thus, for the CG of .488 in Table 1, $t = .375$ ($p > .4$). This clade is effectively symmetrical.

But Is It a Clade?

Before turning to examples, it is important to discuss how clades are distinguished. Earlier, we assumed that each of the five styles of headstones shown in Fig. 1 comprises a (historical) class within a single monophyletic group, or clade, and thus that Fig. 4a represents a clade-diversity diagram. This assumption is incorrect. Dethlefsen and Deetz (1966:503-504), who compiled the data for and published the original version of our Fig. 1, indicate that four of the five types—death's heads, heart-mouth, medusa, cherub—comprise what they term the "slate-gravestone tradition" and imply that the fifth type—urn-and-willow—comprises a different tradition. This change in tradition is "associated with a marked alteration" from headstones with arched shoulders that signify the slate-gravestone tradition to square-shouldered headstones on which urn-and-willow motifs are found (Dethlefsen and Deetz 1966:503). As indicated in Fig. 1, headstones of the latter (and later) tradition replace those of the former (and earlier) tradition between 1800 and 1820. Thus, an analytically correct CG value can be calculated for the slate-gravestone (or arched-shoulder) tradition (CG = .500) but not for the urn-and-willow tradition, as the termination of the

the second, and CG values greater than .5 denote the third. Using random simulations, they determined that statistically significant ($p < .05$) bottom-heavy asymmetry was found only in clades with a CG of less than .428 and top-heavy asymmetry only in clades with a CG of greater than .578 (Kitchell and MacLeod 1988:1192).² In contrast, Gilinsky et al. (1989) argued that one could, using Student's t , determine if differences are statistically significant between a CG value and the mean CG of .500 derived from Kitchell and MacLeod's (1988) simulations. These and additional simulations by others (e.g., Uhen 1996) indicate that the number of time intervals included in a simulation combined with the particular parameters of each simulation (particularly the probability of the appearance of a new variant and the probability that an existing variant will

latter is unknown (it is not specified by Deetz and Dethlefsen).

Their intensive studies of colonial headstones in eastern Massachusetts and temporally associated written documents allowed Deetz and Dethlefsen (1965, 1967, 1971; Dethlefsen and Deetz 1966) to identify the two traditions of headstone production. Would those same traditions be recognizable archaeologically? Perhaps, but perhaps not. The association between shoulder shape and motif is not rigid; Deetz (1977:72) states that the change from cherub to urn-and-willow motifs "is usually accompanied with a change in stone [shoulder] shape" (emphasis added). Shoulder shape could not be used alone to denote a change in tradition. Further, the change in tradition also involved a shift from part-time production of headstones by individuals who had other employment to "a full-time specialty" (Dethlefsen and Deetz 1966:403). This shift in production mode may or may not be visible archaeologically. Finally, Deetz and Dethlefsen (1967, 1971; Dethlefsen and Deetz 1966) used phyletic seriation as well as frequency seriation to help distinguish between the two traditions. The phyletic seriations revealed gradual changes in decorative motifs and thus suggested heritable continuity within the slate-gravestone tradition, but no such continuity between the motifs of that tradition and those of the urn-and-willow tradition were found. Although phyletic seriation may reveal such continuity, we note that this seriation technique demands much finer-scaled classification of the headstones than is indicated in Fig. 1. The phyletic seriation presented by Deetz and Dethlefsen (1967:34-35) of the same material graphed in Fig. 1 and comprising only the slate-gravestone tradition involved 17 styles; the phyletic seriation presented by Deetz (1977:79) of the same material involved 14 styles.

The analytical technique of phyletic seriation is underpinned by a notion of gradual change within a line of heritable continuity

that is readily accounted for by the vagaries of transmission (whether lack of fidelity is intentional and irrespective of the source of new variants). A break in such a line can vary in magnitude. A break of lesser magnitude has typically been interpreted as resulting from either a significant diffusion or enculturation event (if the break is at the scale of one or several of the extant types) or an immigration event [if the break is at the scale of all or nearly all extant types (e.g., Thompson 1956)]. Another criterion used to help distinguish between these two kinds of discontinuities is the presence-absence of styles that could conceivably serve as phyletic ancestors to the new styles. If no such styles are present, then immigration is suggested; if some ancestral styles are present, then diffusion or enculturation is suggested (e.g., Ford 1952; Steward 1929). These criteria are included in the frequency seriation technique used by Lipo et al. (1997) to identify discontinuity in historical sequences that imply a change from one lineage to another.

Our point here is that in order to correctly construct and interpret clade-diversity diagrams, a major analytical effort must comprise the construction of artifact lineages, which in turn rests on classification (e.g., Patterson and Smith 1988). Further, not only is the shape of a clade dependent on classification, the richness and duration of a clade are influenced by the taxonomic level used to quantify clade diversity (Stanley et al. 1981). Failure to expend this effort may result in a clade-diversity diagram such as that shown in Fig. 4b, which, though largely correct, implies by the included dotted line that the urn-and-willow tradition evolved from the slate-headstone tradition, which according to Dethlefsen and Deetz (1966) is incorrect. They had historical documents to help them; another analyst without such documents might have defined a "headstone tradition" and included both the round-shouldered slate-gravestone tradition and the urn-and-willow tradition ma-

terials. Given this definition of the tradition, Fig. 4b would be correct. Clearly, much more thinking regarding how we identify and define traditions is required. Such definition and identification reduces to a classification problem, an arena that in our view has seen far too little thought among archaeologists as yet (O'Brien and Lyman 2000).

EXAMPLES OF MEASURING AND EXPLAINING THE HISTORY OF VARIATION

If one wishes to identify instances of stimulated variation, one must first establish that the sequence of materials under study comprises a lineage, and that requires the use of frequency seriation, percentage stratigraphy, and/or phyletic seriation. Only after it can be shown that the sequence also comprises a line of *heritable continuity* is it reasonable to monitor the history of diversity. The burst of variation signifying stimulated variation must then be distinguished from the random background generation of novel variants, and calculating CG values will help in this endeavor. The CG values that are significantly different in a statistical sense from the random background generation of variation would suggest stimulated variation had occurred; statistically nonsignificant CG values would suggest random background variation or multicausal generation of variation. We now turn to examples of monitoring diversity within artifact clades to illustrate these points and note at the outset that the data sets we discuss are less than ideal for *interpretive purposes*. They do, however, provide excellent examples of some of the analytical hurdles that must be cleared if clade-diversity diagrams are to be of use archaeologically.

Diversity in Radios

As one example of stimulated variation, Schiffer (1996:657) graphed changes in the

frequency of U.S. companies that manufactured vacuum-tube radios for the home market between 1920 and 1955. We assume his data reflect heritable continuity, as radio manufacturers either overtly shared technological information or they covertly reengineered products sold by other manufacturers. Further assuming, as does Schiffer (1996:656), that the number of classes of radios extant during a time period is correlated directly with the number of companies extant during that time period, we graphed Schiffer's data (Table 2) in a form similar to a clade-diversity diagram (Fig. 5). The graph appears to be bottom heavy—there appear to be more companies early and progressively fewer later in time. It appears that many companies—and probably types of radio—arose quickly and then slowly died out. Schiffer's (1996) clade-diversity diagram for manufacturers of vacuum-tube radios (Fig. 5) has a CG value of .443. Student's t is 1.78 ($p < .05$) and indicates that significantly more companies manufactured vacuum-tube radios (and thus, presumably, produced more kinds of such radios) prior to the middle of the time period considered than after that point.

As a second example of stimulated variation, Schiffer (1996:658) discussed the frequencies of "different portable radio models manufactured and sold in the United States" between 1920 and 1955. His data are summarized in Table 2 and are graphed as a clade-diversity diagram in Fig. 6. This diagram has a CG value of .686; statistically, it is significantly top heavy ($t = 5.81$, $p < .01$). Schiffer (1996:657) provides an explanation for the top-heavy nature of this graph, but he knows the particular historical contingencies of the events depicted in the graph. Explaining such changes is fraught with analytical difficulties when the historical contingencies of graphed events are unknown. In other words, not knowing the selective environments in which these events took

TABLE 2

Frequency Data for Vacuum-Tube Radio-Manufacturing Companies and Models of Portable Radios in the United States between 1920 and 1955^a

Year	N of companies manufacturing vacuum-tube radios	N of models of portable radios
1920	4	0
1921	8	1
1922	36	4
1923	99	24
1924	141	22
1925	156	18
1926	105	5
1927	101	4
1928	81	0 [4]
1929	97	4
1930	99	6
1931	105	0 [4.5]
1932	88	3
1933	87	7
1934	87	2
1935	75	2
1936	67	8
1937	61	3
1938	58	4
1939	60	103
1940	58	49
1941	53	55
1942	0 [62.2]	4
1943	0 [71.4]	0 [11.75]
1944	0 [80.6]	0 [19.5]
1945	0 [89.8]	0 [27.25]
1946	99	35
1947	111	75
1948	86	98
1949	57	41
1950	49	69
1951	47	22
1952	45	47
1953	43	54
1954	44	55
1955	39	55

^a Data from Shiffer (1996). Values in brackets were derived by interpolation (see text for discussion).

place could make identifying this graph as an instance of stimulated variation difficult.

A practical point concerning calculation of the CG values needs to be made here. The clade-diversity diagrams in Figs. 5 and 6 contain gaps for the years of World

War II. Schiffer (1996:656) indicates the war "caused [a] hiatus in the manufacture of home radios." We know the reason for this particular gap because we know the precise historical contingencies influencing the manufacture of radios at the time. How might such gaps be dealt with were they to appear in prehistoric materials for which the historical contingencies are unknown? One obvious response to such gaps—whether one is a paleobiologist or an archaeologist—is to blame them on deficient samples. For example, Eldredge and Gould (1977:27) observe that "it is unfortunately true that we cannot be sure that a collection of fossils is a truly representative sample of a biological population." Among paleobiological treatments of clade diversity there is acknowledgment of sample-adequacy issues, and efforts have been made to ascertain if the available samples are influencing measurements of diversity (e.g., Signor 1982).

Gould et al. (1987) accept as a fact that their samples *are* deficient as evidenced by the way they deal with uneven distributions of fossil samples across temporal periods. They caution that if the durations of periods used to construct a clade-diversity diagram vary—say, if the first sample of fossils falls within the first million years of the time period under study, the next sample dates between 7 and 8 million years ago, the next between 10 and 11, and the last between 16 and 17 million years ago—the CG value will be skewed if one simply uses the median age of each sample. Their preferred method of compensating for samples assumed to be deficient is to determine the desired duration of chronological periods and then to derive via interpolation a richness value for each temporal period between those periods with known richness values such that each period is of similar duration.

In the example Gould et al. (1987:1438) describe, seven fossil samples are unevenly distributed over a period of 26 mil-

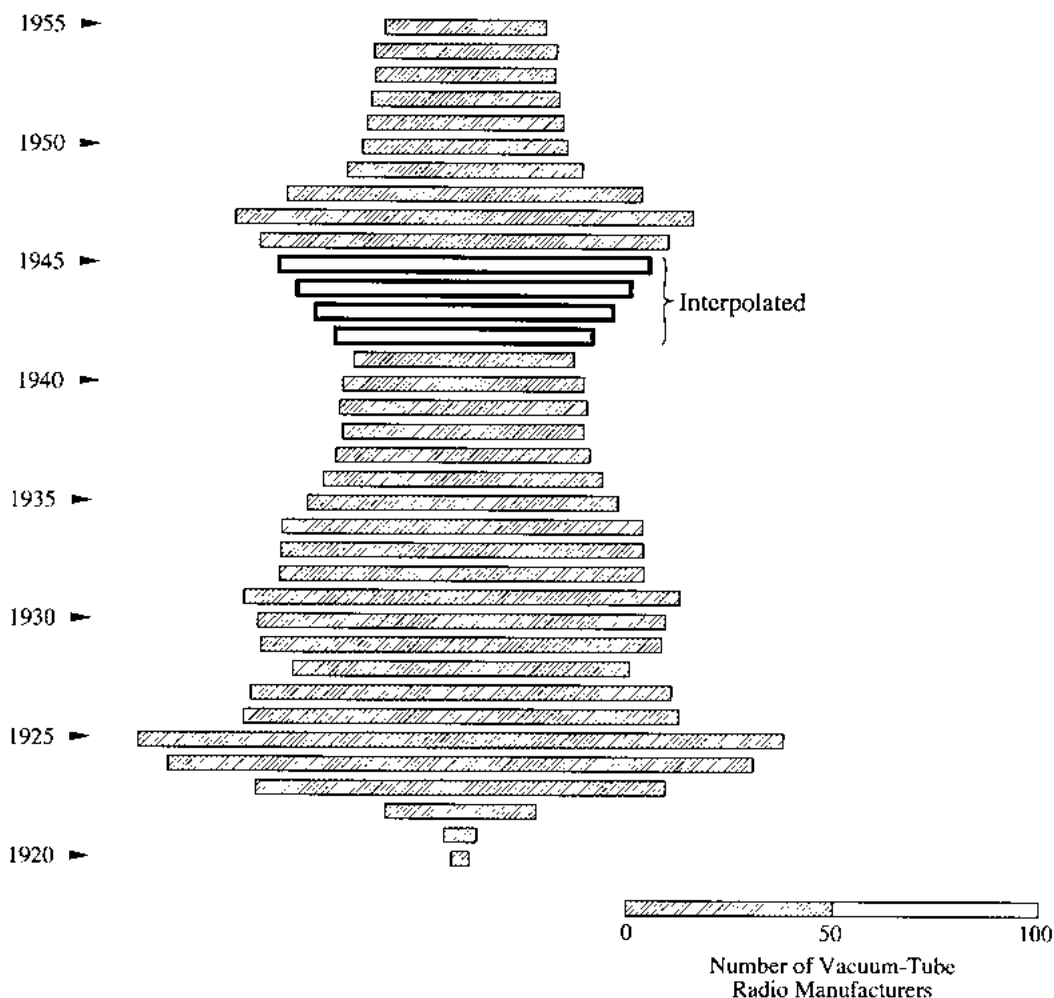


FIG. 5. Clade-diversity diagram of the frequency of companies in the United States that manufactured vacuum-tube radios. Data from Table 2; interpolated values added (see text for discussion).

lion years. Because they desire periods of equal duration to compensate for the effects of varied period duration on the calculation of the CG value, they derive by interpolation 18 richness values plus a zero at both ends of 26 million years (7 observed values, 18 interpolated values, and 2 zeros at either end), such that a value is available for year zero and every subsequent 1 million years of the 26-million-year span. All richness values—observed or interpolated—are then included

in the calculation of the CG value. Gould et al. (1987:1439) favor this method because it is, in their view, the "least biased" of the three they consider. They note that in the example they discuss, the CG value is .516 ($t = .5, p > .5$) without interpolated richness values and .448 ($t = 1.624, p > .05$) with the latter values included. The addition of the interpolated values does not change the CG value in any statistically significant manner, though Gould et al. (1987) find the difference important and

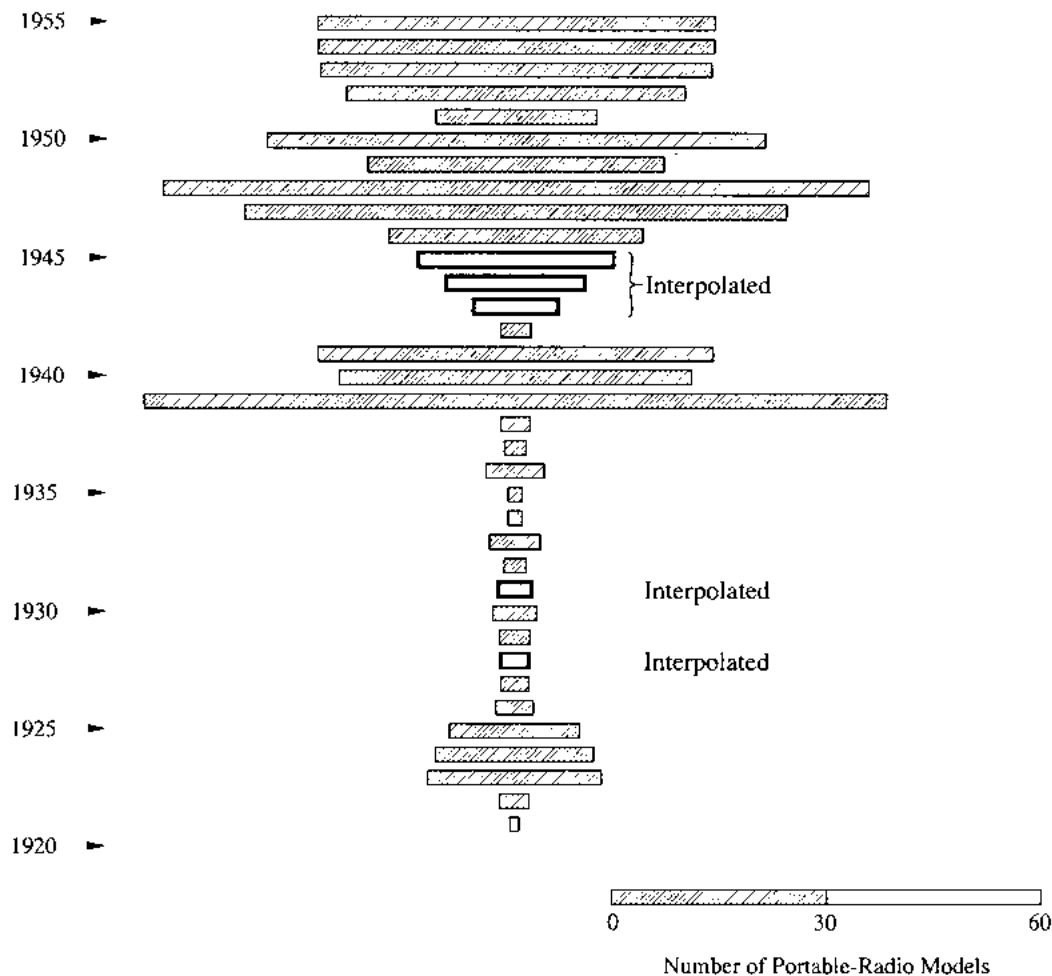


FIG. 6. Clade-diversity diagram of models of portable radios manufactured in the United States. Data from Table 2; interpolated values added (see text for discussion).

alter their interpretation of the graph from being more or less symmetrical to being bottom heavy.

One might well ask why a best-fit polynomial equation is not used to determine expected values for all temporal points rather than just those with richness values of zero—a procedure suggested in archaeology (e.g., Kintigh 1984; McCartney and Glass 1990). But were one to pose this question, we suspect that it would quickly become obvious that even the interpolation procedure favored by Gould et al. (1987) presumes at least in part precisely what

clade-diversity analysis is attempting to determine—fluctuation in the richness of types over time. As Rhode (1988:711) observes, such an analytical step implies “that a great deal already is known about the nature of the archaeological record.” That is, one is presuming—via the interpolation process—exactly what one is attempting to determine—*changes* in diversity over time. Precisely this point has recently been recognized by paleobiologists (Kirchner and Weil 1998).

We calculated CG values for the radio data in Table 2 after substituting interpo-

lated richness values for all zero values. We refer to these as CG_{int} to indicate that interpolated richness values are included in their calculation. The interpolated values are listed in Table 2 and are shown in Figs. 5 and 6. The CG_{int} value for Fig. 5 is .468 ($t = .9995, p > .1$) and the CG value is .443 ($t = 1.78, p < .05$); the CG_{int} value for Fig. 6 is .682 ($t = 5.685, p < .01$) and the CG value is .686 ($t = 5.81, p < .01$). Richness values derived by interpolation do not create statistically significant changes in the CG values for models of portable radios (Fig. 6), but they do result in such changes for vacuum-tube radio manufacturers (Fig. 5). Given our cautions in the preceding paragraph, we suggest calculating both CG and CG_{int} values may be advisable.

On the one hand, there does not seem to be a statistically significant difference between the history of vacuum-tube radio diversity and randomly generated background variation within a clade ($p > .1$ for CG_{int} , but $p < .05$ for CG). Although Schiffer's research indicates that the Fig. 5 graph in fact comprises an instance of stimulated variation, the statistical test of CG_{int} alone would not allow us to identify that graph as such. Statistical tests of the clade-diversity diagram in Fig. 6, on the other hand, suggest that stimulated variation might be represented ($p < .05$ for both CG and CG_{int}). We say "might" for three reasons. First, with respect to Fig. 5, recall Gilinsky and Bambach's (1986) remark: Perhaps so many diverse and somewhat conflicting "causal factors" are included that any statistical indication of stimulated variation is muted. Second, with respect to Fig. 6, the CG values indicate the graph is top heavy. It is unclear in Schiffer's (1996) discussion if stimulated variation will variously produce top-heavy, bottom-heavy, or both kinds of clade-diversity diagrams. We suspect it could produce both. Third, the diagrams in Figs. 5 and 6 lack their tops, and thus we do not know what the complete history of diversity for either clade was like.

Any comparison to a random model and any effort to calculate a CG value is spurious because that value will automatically suggest a top-heavy graph (Gould et al. 1977, 1987). As a result, Schiffer's empirical cases provide statistically ambiguous indications of stimulated variation, which is *not* to say that they do not comprise examples of stimulated variation.

The preceding corroborates our earlier contention that identifying instances of stimulated variation in the prehistoric record will have steep data requirements. That both stylistic and functional variation are probably plotted in Figs. 5 and 6 no doubt exacerbates analytical and interpretive ambiguity. We suspect that were these two kinds of variation distinguished and plotted separately, rather different results would be produced. This can be shown by turning first to an example in which the kinds of variation under study are better known and then to an example in which only stylistic variation is considered.

Diversity in Great Basin Projectile Points

The history of types of projectile points from Gatecliff Shelter in Nevada (Thomas and Bierwirth 1983) is presented in Fig. 7; observed richness and chronological assignments (after Thomas 1983a:174) are summarized in Table 3. A clade-diversity diagram for these materials is shown in Fig. 8. Only two richness values—for Horizons 11 and 13—require interpolation. The CG_{int} value was calculated by assigning each assemblage of projectile points to the midpoint of its temporal range—a procedure that Gould et al. (1987:1439) indicate will put "too much weight on [temporally] more closely spaced samples." However, if the duration of Horizon 2 is assumed to be 50 years and the durations of Horizons 4, 5, and 6 are assumed to be 650 years each, then the duration of the horizons is not correlated with the sequence of horizons (Spearman's $\rho = .11$,

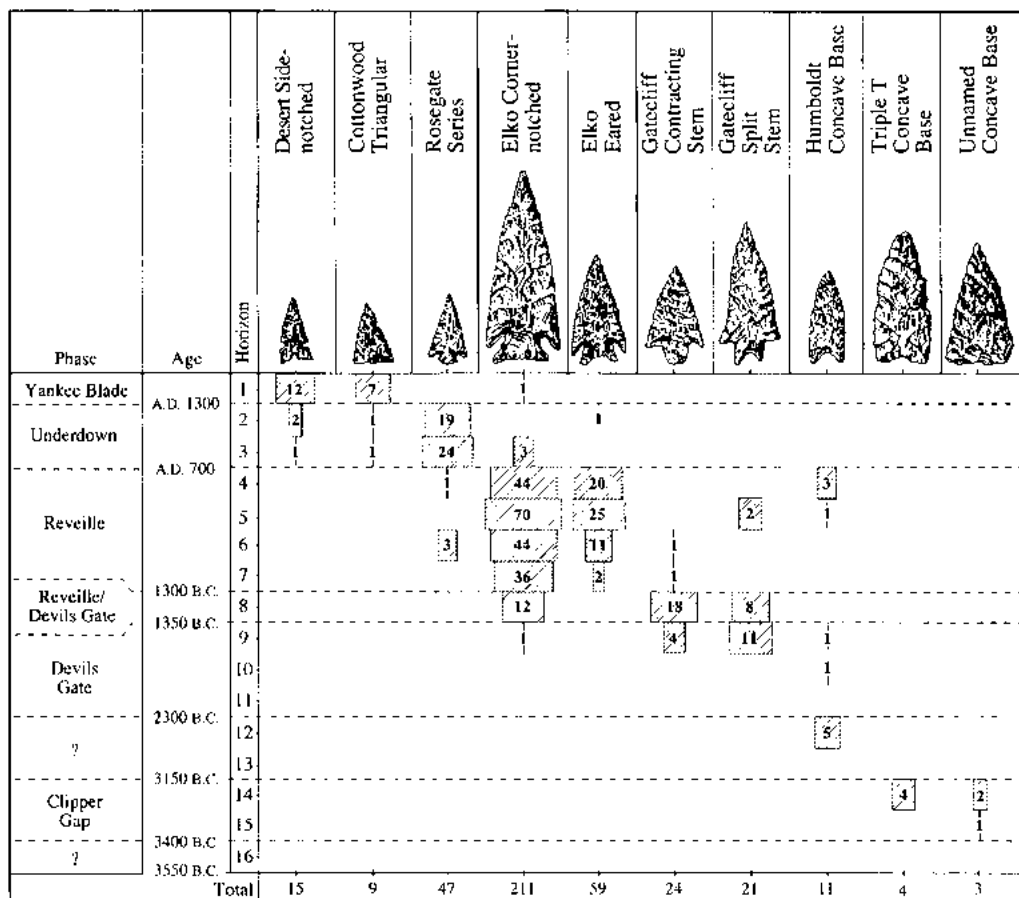


FIG. 7. Percentage-stratigraphy graph of Gatecliff Shelter, Nevada, projectile points (after Thomas and Bierwirth 1983).

$p > .6$). Further, lumping Horizons 4–6 into a single unit 1950 years in duration and assigning Horizon 2 a duration of 1 year, as is implied in Table 3, the correlation between horizon duration and horizon sequence is no different from zero ($\rho = .01$, $p > .9$). Gould et al. (1987:1439) indicate such a result suggests period duration will not influence CG values.³ The CG_{int} value for the clade-diversity diagram in Fig. 8 is .5995 ($t = 3.108$, $p < .001$) and indicates that the diversity of projectile-point types increases through time.

In the following, we consider the clade-diversity diagram in Fig. 8 and its associated CG_{int} value. It is important, therefore,

to point out that were we to follow Gould et al.'s (1987) preferred procedure of designating temporal periods of equal duration, those periods each might be of 100 years duration, as that is the smallest temporal increment that can be extracted from Table 3. That procedure results in many more interpolated than observed richness values, and thus we do not pursue it here except to note that it, too, tends to produce a top-heavy clade. Alternatively, we might lump various horizons in an effort to produce collections of more or less equal duration. One possible result would be: Horizons 16–13 (500 years, two types), Horizon 12 (750 years, one type),

TABLE 3
Temporal Distribution of Projectile-Point Richness
in Gatecliff Shelter, Nevada

Horizon	Age	Observed richness (interpolated)	Scaled time
—	post 1500	0	1.0
1	1300-1500	3	0.9375
2	1300	4	0.875
3	700-1300	4	0.8125
4*	50 A.D.	4	0.750
5*	—	4	0.6875
6*	1250 B.C.	4	0.625
7	1300-1250	3	0.5625
8	1350-1300	3	0.5
9	1450-1350	4	0.4375
10	2100-1450	1	0.375
11	2300-2150	0 (1)	0.3125
12	3050-2300	1	0.25
13	3150-3050	0 (1.5)	0.1875
14	3300-3150	2	0.125
15	3400-3300	1	0.0625
16	3550-3400	0	0.0

* Horizons 4, 5, and 6 span the period 1250 B.C.—50 A.D.

Horizons 11–10 (850 years, one type), Horizons 9–6 (850 years, six types), Horizon 5 (650 years, four types), Horizon 4 (650 years, four types), and Horizons 3–1 (800 years, five types). This procedure, too, produces a rather top-heavy clade. Thus, we find the clade illustrated in Fig. 8 satisfactory for discussion purposes. What can we make of it?

First, it is important to note that Thomas (1983b:425–431) documents in detail that the richness of stone-tool types per cultural horizon at Gatecliff Shelter is a function of sample size—larger samples produce more types. Sample size and observed projectile-point richness data summarized in Fig. 7 are correlated (Pearson's $r = .744$, $p < .001$), suggesting that caution would be warranted were one to interpret the CG value derived from those data. For sake of discussion here, we assume the effects of sample size are insignificant in the following discussion of the CG_{int} value. We further assume that the

clade-diversity diagram (Fig. 8) is not truncated at either end. Granting these two assumptions allows insights to the history of Great Basin projectile-point diversity that can be phrased as hypotheses to be tested with additional data.

Each projectile-point type displays a more or less unimodal frequency distribution when plotted against vertical provenience, but those distributions are imperfect (Fig. 7). Various explanations could be mustered to account for this kind of distribution, including stratigraphic mixing (e.g., Burgh 1959) and recycling and reworking (e.g., Flenniken and Wilke 1989). Most recently, it has been suggested that at least some of the variation in the distributions of the types is attributable to functional causes (Beck 1995, 1998). None of these arguments invalidates the implied temporal ordering of the types, as they rather consistently fall in the indicated temporal order (e.g., Bettinger et al. 1991; O'Connell and Inoway 1994). But these arguments do underscore the fact that some of the definitive attributes of the types measure time—are potentially stylistic—and some attributes measure functional change (see also Hughes 1998). The two sets of attributes need not be and likely are not mutually exclusive simply because they are both found on the same point. It is likely that the point "types" at Gatecliff Shelter are combinations of both kinds of attributes and hence produce less than perfect unimodal frequency distributions. Beck (1995) found that side notching occurred frequently early (on large points), then dropped in frequency, and finally became more frequent again late in time (on small points). This is not unusual for a functional trait. She also noted that the proximal shoulder angle of all points was the single attribute that correlated significantly with time, as a stylistic attribute should, though this does not demonstrate that the proximal shoulder angle actually is stylistic because functional traits can also be transmitted or

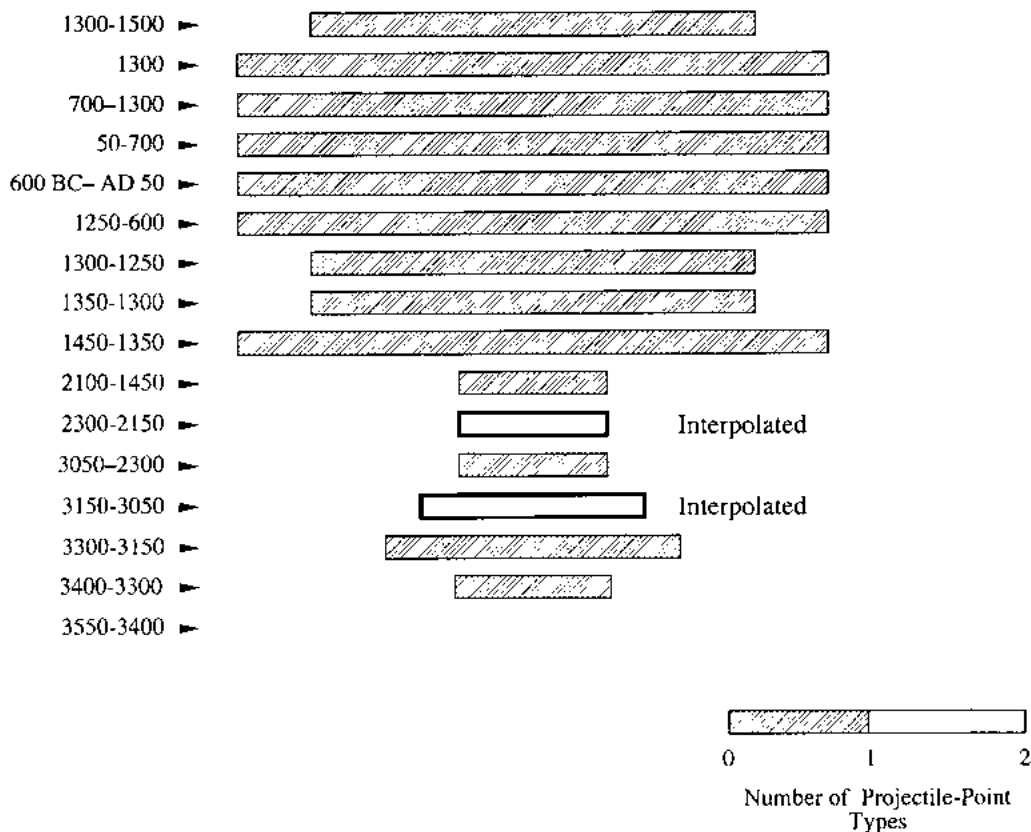


FIG. 8. Clade-diversity diagram for Gatecliff Shelter projectile points. Data from Table 3; interpolated values added (see text for discussion).

inherited and thus may correlate with time's passage.

Others have suggested that change in the delivery system—lance or javelin to atlatl dart to bow and arrow—was the driving selective force resulting in change in Great Basin projectile points (e.g., Beck 1995; Musil 1988; see also Hughes 1998). Recalling that the t statistic for CG_{m1} is significant ($p < .01$), change in the diversity of Gatecliff Shelter points might, then, comprise an instance of stimulated variation. We say "might" because of the kinds of variation measured by the projectile-point types; some of the attributes appear to be functional and others stylistic. Further, some of the definitive attributes of the Gatecliff Shelter projectile-point types

may have been merely sorted rather than directly subject to selection. That is, some of the attributes of the hafting elements used to designate point types changed as the engineering/design requirements for efficient hafting of a point on a lance or javelin changed to those for efficient hafting of a point on an atlatl dart and then to those for efficient hafting of a point on an arrow (Hughes 1998). We propose, then, that the increasing diversity of projectile-point types comprises an instance not of stimulated variation per se but rather an instance of an increase in the number of sets of design constraints that resulted from an increase in the number of weapon delivery systems.

What is design constraint? "Constraint is

a theory-bound term for causes of change and evolutionary direction by principles and forces outside an explanatory orthodoxy" (Gould 1989:519; see also Carroll 1986; Stearns 1986). The suggestion that change in the delivery system of projectile points was the selective force resulting in change in projectile-point form is orthodox functionalism (in Gould's and Stearn's terms). But note that minimally two scale shifts are made when one calls on weapon delivery systems as a functional explanation for change in the attributes of projectile-point haft elements. The scale shifts are from the complete weapon to the point (type) and from the point type to the attributes of the point, specifically those attributes of the hafting element. The particular history of delivery systems drove the increase in diversity of the attributes of projectile-point hafting elements. Each weapon delivery system posed a set of unique engineering design requirements—constraints—on the sorts of hafting element that would work efficiently; as new delivery systems were added, new constraints on projectile-point haft elements came into play. In other words, attributes of haft elements were sorted as a result of design constraints; they were not the direct focus of selection (e.g., Vrba 1989). That honor, in light of the preceding, seems to reside with the total weapon delivery system, a much larger scale entity than the scale at which projectile-point types are distinguished.

This is not the place to delve further into this issue. Our discussion is meant merely to outline the kinds of factors that must be considered if one wishes to explain the history of novel variants—their appearance, replication, and disappearance—and to identify instances of stimulated variation. So far, we have used Schiffer's radio data and data on Great Basin projectile points to explore this issue, but these data appear to comprise both stylistic and functional variation. We turn next to data comprising only stylistic variation.

Lower Mississippi Valley Pottery

The "seriation graph" produced by Phillips et al. (1951:Fig. 21) for ceramic sherds from the St. Francis River area of the Lower Mississippi Alluvial Valley provides an example of changes in the richness of stylistic variants. Because that graph was originally published as a large foldout, we reproduce only a representative portion of it in Fig. 9. The complete graph, founded in part on superposed collections and in part on surface collections, presents an ordering of 58 assemblages. The 47,413 sherds included in the arrangement represent 18 types and span five "periods." Phillips et al. (1951) used surface treatment/decoration as the basis for many of their types. There were few types for the early periods and more for the later periods—a reflection of the increasing variation in decoration in the later periods. The "seriation graph" thus appears to be top heavy, as does the clade-diversity diagram (Fig. 10) derived from it; the CG statistic of .575 ($t = 2.343$, $p < .01$) calculated from clade-diversity data confirms this appearance.

Not only does the richness of types increase through time in these data, so too does sample size (Table 4). These two variables are correlated when all 58 assemblages are included (Pearson's $r = .69$, $p < .002$); the two variables also are correlated when samples are lumped by chronological period (Pearson's $r = .93$, $p < .05$). These statistics suggest that caution may be warranted when interpreting the history of ceramic diversity represented by the data. But, as with the Gatecliff Shelter projectile points, for sake of discussion we assume sample-size effects are insignificant. We also assume that the ends of the clade are not truncated. The temporal scale of the "seriation graph" is ordinal, and thus we do not know how the assemblages vary in duration. But recall that one requirement of the seriation method is

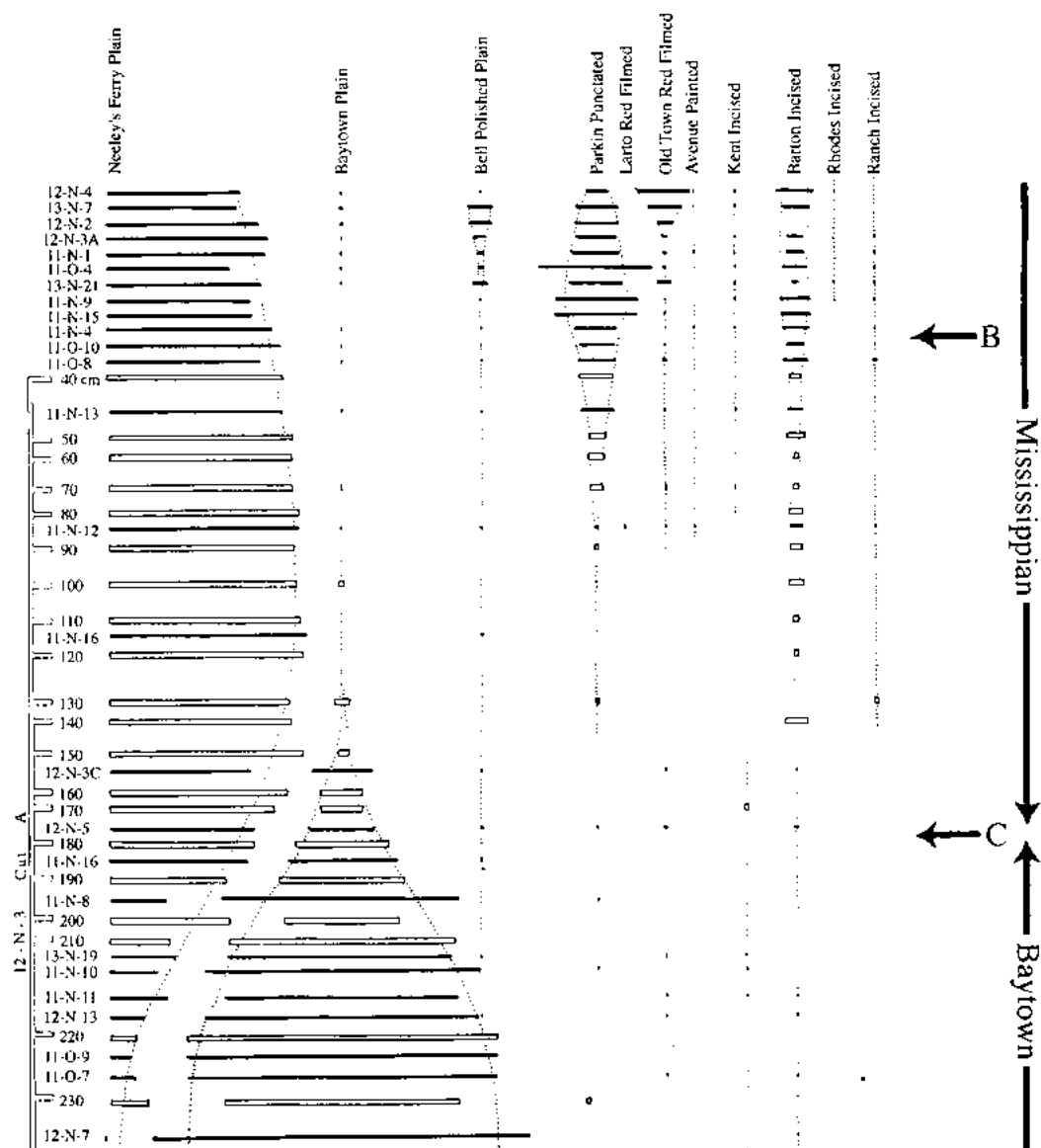


FIG. 9. Chronological ordering of ceramic sherd assemblages from excavated and surface-collected sites in the St. Francis River area of the Lower Mississippi Alluvial Valley. The ordering was based primarily on pottery types from excavated levels at site 12-N-3 (denoted by the E-shaped figure on the left margin). Surface-collected assemblages were placed in the chronology based on percentages of types denoted by bar width. Only 46 of the total 58 assemblages are shown; these represent the late portion of the chronology. Nine rarely represented types are not included. Boundaries between "periods" are denoted by bold horizontal arrows labeled with capital letters (after Phillips et al. 1951).

that the assemblages seriated must be of equivalent duration. The "seriation graph" suggests this requirement is at least approximated, else the frequency

distributions of the graphed types would not approximate unimodal curves (Fig. 9). Therefore, we assume that the temporal duration of each assemblage is more or

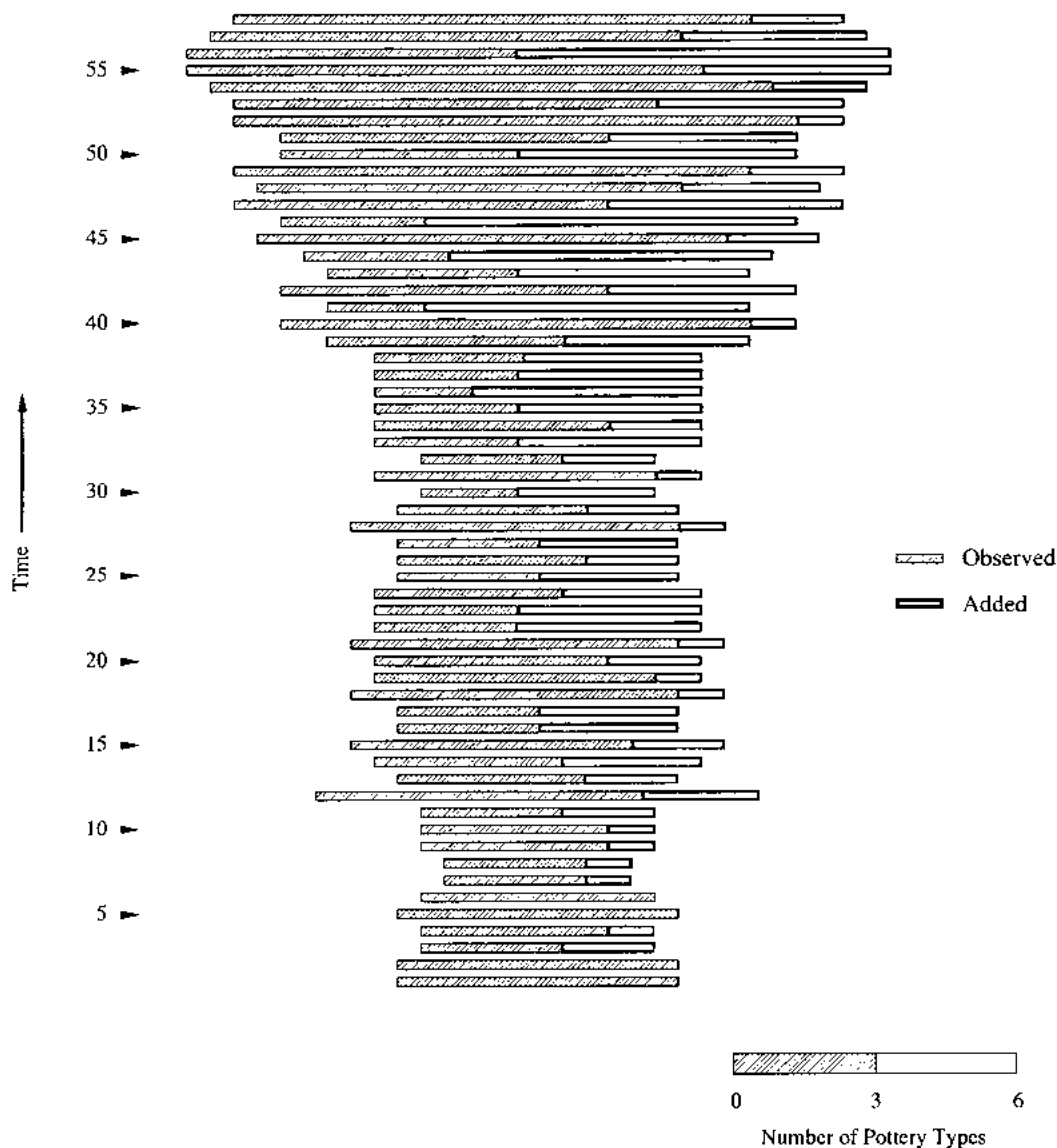


FIG. 10. Clade-diversity diagram for St. Francis River area ceramic types. "Added" richness values are based on Phillips et al.'s (1951) beliefs regarding the temporal span of types (see Fig. 9 and text for discussion). Data from Table 4.

less equivalent to the duration of every other assemblage.

The "seriation graph" for the St. Francis River area represents Phillips et al.'s beliefs regarding the temporal continuity of various pottery types that occur discontinuously in this particular empirical record.

Their beliefs are signified in the graph by dotted lines connecting various samples within individual type columns (see, for example, the column under the type "Bell Polished Plain" in Fig. 9). Tallying richness values for each assemblage according to the combined observed types as

TABLE 4
Observed Ceramic Type Richness, Number of Additional Ceramic Types Believed to be Present,
and Sample Size for St. Francis River Area, Lower Mississippi Alluvial Valley

Assemblage	Observed-type richness	Number of additional types believed present	Sample size	Scaled time
—	0	0	—	1.000
58	11	2	731	0.983
57	10	4	674	0.966
56	7	8	254	0.949
55	11	4	1416	0.932
54	12	2	6898	0.915
53	9	4	722	0.898
52	12	1	2030	0.881
51	7	4	2590	0.864
50	5	6	686	0.847
49	11	2	7287	0.831
48	9	3	4723	0.814
47	8	5	151	0.797
46	3	8	142	0.780
45	10	2	4925	0.763
44	3	7	117	0.746
43	4	5	74	0.729
42	7	4	164	0.712
41	2	7	134	0.695
40	10	1	3063	0.678
39	5	4	106	0.661
38	3	4	65	0.644
37	3	4	60	0.627
36	2	5	103	0.610
35	3	4	113	0.593
34	5	2	99	0.576
33	3	4	59	0.559
32	3	2	83	0.542
31	6	1	273	0.525
30	2	3	73	0.508
29	4	2	69	0.492
28	7	1	698	0.475
27	3	3	56	0.458
26	4	2	193	0.441
25	3	3	87	0.424
24	4	3	179	0.407
23	3	4	103	0.390
22	3	4	146	0.373
21	7	1	266	0.356
20	5	2	211	0.339
19	6	1	298	0.322
18	7	1	1187	0.305
17	3	3	127	0.288
16	3	3	71	0.271
15	6	2	1478	0.254
14	4	3	98	0.237
13	4	2	488	0.220
12	7	2	117	0.203
11	3	2	106	0.186

TABLE 4—Continued

Assemblage	Observed-type richness	Number of additional types believed present	Sample size	Scaled time
10	4	1	109	0.169
9	4	1	223	0.153
8	3	1	158	0.136
7	3	1	47	0.119
6	5	0	629	0.102
5	6	0	987	0.085
4	4	1	181	0.068
3	3	2	285	0.051
2	6	0	713	0.034
1	6	0	288	0.017

well as the types believed to be present—what we term, somewhat inaccurately, interpolated presences—increases the richness values of 54 of the 58 assemblages. The four assemblages with unaltered richness values are among the six oldest assemblages in the ordering (Table 4). The richness of each of the oldest 27 assemblages increases by an average of 1.8, whereas the richness of each of the 31 youngest assemblages increased by an average of 3.7. This suggests that the CG value of .575 may increase when calculated as a CG_{int} value; the latter is .589 ($t = 2.624$, $p < .005$). Both values indicate a top-heavy clade, though the latter is a bit more top heavy than the former.

Richness increases via interpolation in virtually all assemblages throughout the period graphed, but interpolated richness values per assemblage increase progressively more rapidly as assemblages become younger in age, as in Fig. 11, thereby increasing (making more recent in time) the center of gravity of the clade-diversity diagram when it includes interpolated values. The net result is that the CG value is less than the CG_{int} value. The small difference in the CG and CG_{int} values suggests to us that Phillips et al. (1951) knew a great deal about the ceramic chronology of the Lower Mississippi Alluvial Valley. It was precisely this additional knowledge

that informed their adding dotted lines to various of the type columns—what we termed interpolations. Their proposed chronology has stood the test of additional data and analyses. What, then, can we make of the increasing richness of ceramic types?

Because attributes of decoration form the basis of the types, we believe the types are stylistic. The relatively smooth unimodal frequency distributions of the types lend credence to this belief but do not confirm it. Later analyses of Phillips et al. (1951) types suggest that they are, in fact, largely stylistic (Lipo et al. 1997). Why, then, does the richness of types increase late in time, producing a top-heavy clade? We find Braun's (1995) and Lipo et al.'s (1997) discussions particularly helpful. The richness of decorative types or styles is correlated with social dynamics, particularly the frequency of group interaction or intergroup transmission (Braun 1995). Decreased intergroup transmission and increased intragroup transmission result in increased localization of decorative styles (Lipo et al. 1997), much like peripatric speciation results from decreased gene flow between reproductively isolated populations. Increasing stylistic diversity or richness may well reflect, then, a decrease in the panmictic characteristics of the human populations in the St. Francis area.⁴

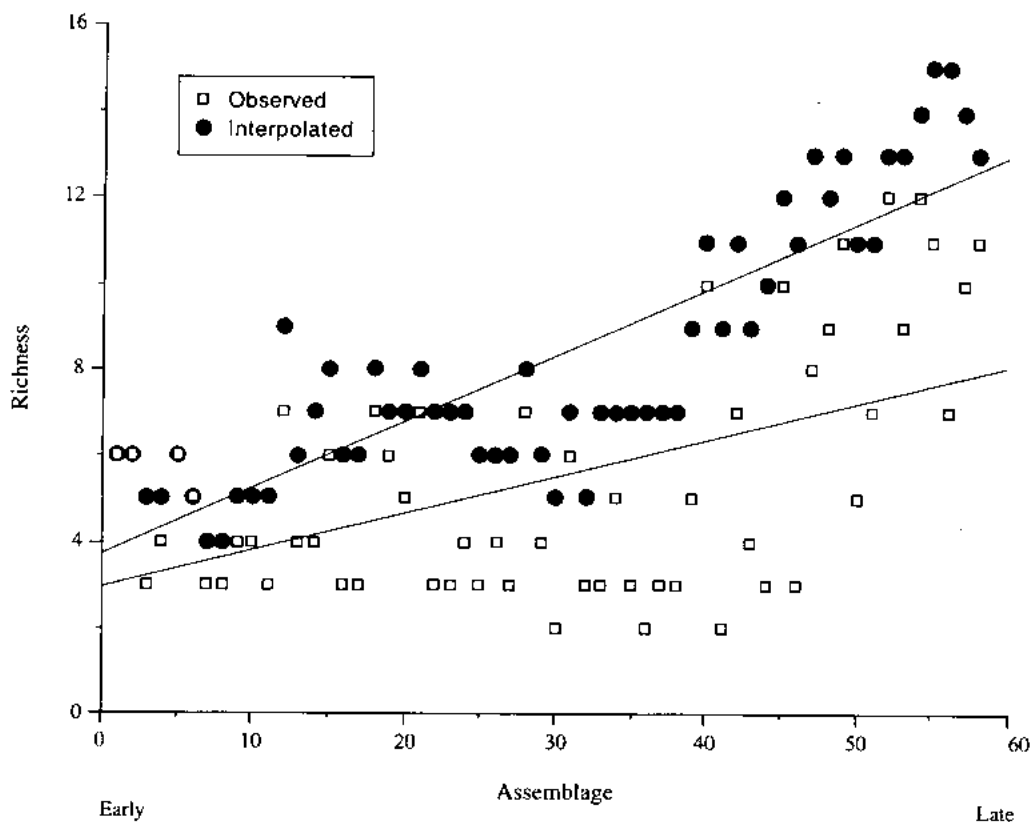


FIG. 11. Changes in richness of pottery types across 58 assemblages from the St. Francis River area; data from Table 4. The lower, less steep line is the simple best-fit regression line for the data plotted as squares (observed richness values; Pearson's $r = 0.51$). The upper, steeper line is the simple best-fit regression line for data plotted as dots [interpolated (observed plus added) richness values; Pearson's $r = 0.85$].

That is, potters became, for whatever reason, more isolated and their decorative styles evolved in independent directions, resulting in progressively more types. Lipo et al. (1997) have recently indicated how these local areas might be identified archaeologically.

Is the increase in diversity of ceramic styles in the St. Francis River area an instance of stimulated variation? Perhaps it is. Increasing localization of decorative styles may result as symbols of one's social identity become more important; ceramic decorative styles transmit information regarding the pottery user's social identity. They also have a production cost,

and thus are "driven by selection" (Lipo et al. 1997:318). Production of progressively more styles over a relatively large area, then, must have been in some sense selected for. Identifying those selective forces is beyond the scope of this article, though we note that some interesting hypotheses have been outlined that might prove applicable (e.g., Braun 1985, 1991; Braun and Plog 1982). Our point here has been to identify the historical shape of artifact-clade diversity as a springboard to detecting when such hypotheses warrant testing. And we emphasize that while the examples we review concern dynamic histories of diversity, a history displaying

stasis in diversity over time also requires explanation.

DISCUSSION AND CONCLUSION

Regardless of the paradigm under which one operates, modern archaeology has as its central focus variation observed in the archaeological record. Explanations of the differential origin and persistence of artifact variants across geographic space and over time vary because of different ontologies, epistemologies, and interpretive algorithms. Yet we agree with the remarks of Schiffer and Dunnell with which we began this article. Many archaeologists seem content to merely measure variation and to append some accommodative interpretations of what they have measured, perhaps with some cautions about the measure of variation used or the influence of the available sample on the measure. Little effort seems to be devoted to building theory to explain the appearance of novel variants and their differential persistence. We have outlined a method for (a) distinguishing between the generation of variation by stimulated variation and the random background generation of variation and for (b) measuring the history of variation, but we emphasize that the method is founded in and derived from a particular theory—Darwinian evolution. The method comprises the construction and analysis of clade-diversity diagrams. The procedure for building the diagrams requires that the units graphed be phylogenetically related so as to reflect a lineage. This in turn requires that frequency seriation or percentage stratigraphy be used to insure that a line of heritable continuity rather than a simple sequence is under study.

Darwinian theory also provides an explanation for the appearance of novel variants and their differential persistence. Stylistic forms persist, or not, simply because of the vagaries of transmission.

Functional forms persist, or not, more as a result of selection than of transmission, though the latter also plays a role. But the distinction between the two kinds of forms is fundamental to monitoring and explaining the history of variation. Thus we find Schiffer's notion of stimulated variation an intriguing one, and we believe that the construction of clade-diversity diagrams, when used in concert with their attendant theory—Darwinism—provides a means to explore and explain the history of artifact diversity within an artifact lineage. The groundwork already laid by paleobiologists provides a way to determine when a clade-diversity diagram diverges significantly from a randomly generated pattern and constitutes an instance of stimulated variation. This in turn allows one to begin to explain why some artifact clade-diversity diagrams have the particular shapes they do.

ACKNOWLEDGMENTS

We thank J. Rensberger for providing a copy of his unpublished paper, D. Glover for producing the figures, and M. B. Schiffer for providing copies of his data on radios and radio-manufacturing companies. J. Darwent, R. D. Leonard, E. J. O'Brien, M. B. Schiffer, and S. Wolverton provided helpful comments on several drafts of the article. We also thank J. O'Shea, Carl Lipo, and Bert Voorrips for providing cogent, accurate criticisms.

NOTES

¹ Care must be exercised in building a randomly generated clade-diversity diagram. Many machine-generated sets of random numbers do not comprise truly random sequences of numbers, and thus the shape of so-called randomly generated clade-diversity diagrams are not, in fact, random (Rensberger and Barnosky n.d.). Further, one must be abundantly clear about what is meant by "random," as it might not always mean what a statistician means when used by a biologist or archaeologist (Bookstein 1987).

² The statistically significant values indicated were derived from a simulation of the history of 1000 clades over 63 time intervals. Other simulations with similar parameters produce similar mean CG values and statistically significant minimum and maximum

values [see Kitchell and MacLeod (1988) for details]. Uhen (1996) later derived similar results in an independent set of simulations.

³ As Foote (1991:116) makes clear, the center of gravity of a clade-diversity diagram "inherent in the time scale" may not be .5 simply because of variation in the duration of the periods of which it is comprised. Given the time scale—the set of temporal units and their durations denoted by the cultural horizons—for the Gatecliff Shelter projectile points, the CG inherent in the time scale is .436. This value is calculated as the proportion of scaled time since the initiation of the clade indicated by the median shared temporal boundary between horizons (or periods). Thus, the median boundary is that between Horizons 8 and 9; it falls at 1350 B.C., which is .436 of the total time since the clade's initiation at 3550 B.C. (with the clade terminating at 1500 A.D.). We ignore this problem here and note that were this inherent CG used for statistical comparison rather than .5, the clade-diversity diagram for Gatecliff Shelter projectile points would still be significantly top heavy.

⁴ Dethlefsen and Deetz (1966:508) state that it "appears that [novel variation] is primarily initiated by a small segment of the population and then spreads to the majority." This is an effective statement of what is known in biological evolution variously as geographic, allopatric, or peripatric speciation.

REFERENCES CITED

- Basalla, G.
1988 *The evolution of technology*. Cambridge University Press, Cambridge.
- Beals, R. L., G. W. Brainerd, and W. Smith
1945 Archaeological studies in northeast Arizona. *University of California, Publications in American Archaeology and Ethnology* 44:1-236.
- Beck, C.
1995 Functional analysis and the differential persistence of Great Basin dart forms. *Journal of California and Great Basin Anthropology* 17: 222-243.
1998 Projectile point types as valid chronological units. In *Unit issues in archaeology: Measuring time, space, and material*, edited by A. F. Ramenofsky and A. Steffen, pp. 21-40. University of Utah Press, Salt Lake City.
- Benton, M. J.
1995 Diversification and extinction in the history of life. *Science* 268:52-58.
- Bettinger, R. L., J. F. O'Connell, and D. H. Thomas
1991 Projectile points as time markers in the Great Basin. *American Anthropologist* 93:166-172.
- Binford, L. R.
1965 Archaeological systematics and the study of culture process. *American Antiquity* 31:203-210.
1968 Archeological perspectives. In *New perspectives in archeology*, edited by S. R. Binford and L. R. Binford, pp. 5-32. Aldine, New York.
- Bookstein, F. L.
1987 Random walk and the existence of evolutionary rates. *Paleobiology* 13:446-464.
- Braun, D. P.
1985 Ceramic decorative diversity and Illinois Woodland regional integration. In *Decoding prehistoric ceramics*, edited by B. A. Nelson, pp. 128-153. Southern Illinois University Press, Carbondale.
1991 Why decorate a pot? Midwestern household pottery, 200 B.C.-A.D. 600. *Journal of Anthropological Archaeology* 10:360-397.
1995 Style, selection, and historicity. In *Style, society, and person: Archaeological and ethnological perspectives*, edited by C. Carr and J. E. Neitzel, pp. 123-141. Plenum, New York.
- Braun, D. P., and S. Plog
1982 Evolution of "tribal" social networks: Theory and prehistoric North American evidence. *American Antiquity* 47:504-525.
- Burgh, R. F.
1959 Ceramic profiles in the western mound at Awatovi, northeastern Arizona. *American Antiquity* 25:184-202.
- Carroll, R. L.
1986 Physical and biological constraints on the pattern of vertebrate evolution. *Geoscience Canada* 13:85-90.
- Cowgill, G. L.
1972 Models, methods and techniques for seriation. In *Models in archaeology*, edited by D. L. Clarke, pp. 381-424. Methuen, London.
- Darwin, C.
1859 *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. Murray, London.
- Deetz, J.
1977 *In small things forgotten: The archeology of early American life*. Anchor, Garden City, NY.
- Deetz, J., and E. Dethlefsen
1965 The Doppler effect and archaeology: A consideration of the spatial aspects of seriation. *Southwestern Journal of Anthropology* 21:196-206.
1967 Death's head, cherub, urn and willow. *Natural History* 76(3):29-37.

- 1971 Some social aspects of New England colonial mortuary art. In *Approaches to the social dimensions of mortuary practices*, edited by J. A. Brown, pp. 30–38. Society for American Archaeology Memoir No. 25.
- Dethlefsen, F., and J. Deetz
1966 Death's heads, cherubs, and willow trees: Experimental archaeology in colonial cemeteries. *American Antiquity* 31:502–510.
- Dunnell, R. C.
1970 Seriation method and its evaluation. *American Antiquity* 35:305–319.
1978 Style and function: A fundamental dichotomy. *American Antiquity* 43:192–202.
1980 Evolutionary theory and archaeology. *Advances in Archaeological Method and Theory* 3:35–99.
1981 Seriation, groups, and measurements. In *Manejo de datos y metodos matematicos de arqueologia*, compiled by G. L. Cowgill, R. Whallon, and B. S. Ottaway, pp. 67–90. Union Internacional de Ciencias Prehistoricas y Protohistoricas, Mexico D.F.
1989 Diversity in archaeology: A group of measures in search of application? In *Quantifying diversity in archaeology*, edited by R. D. Leonard and G. T. Jones, pp. 142–149. Cambridge University Press, Cambridge.
- Eldredge, N., and S. J. Gould
1977 Evolutionary models and biostratigraphic strategies. In *Concepts and methods of biostratigraphy*, edited by E. G. Kauffman and J. E. Hazel, pp. 25–40. Dowden, Hutchinson & Ross, Stroudsburg, PA.
- Evans, J.
1850 On the date of British coins. *The Numismatic Chronicle and Journal of the Numismatic Society* 12:127–137.
- Flenniken, J. J., and P. J. Wilke
1989 Typology, technology, and chronology of Great Basin dart points. *American Anthropologist* 91:149–185.
- Foote, M.
1991 Morphological and taxonomic diversity in a clade's history: The blastoid record and stochastic simulations. *University of Michigan, Contributions from the Museum of Paleontology* 28:101–140.
- Ford, J. A.
1938 *An examination of some theories and methods of ceramic analysis*. Unpublished M.A. thesis, University of Michigan, Ann Arbor.
1949 Cultural dating of prehistoric sites in Virú Valley, Peru. *American Museum of Natural History, Anthropological Papers* 43:29–89.
- 1952 Measurements of some prehistoric design developments in the southeastern states. *American Museum of Natural History, Anthropological Papers* 44:313–384.
- 1962 A quantitative method for deriving cultural chronology. *Pan American Union, Technical Bulletin* No. 1.
- Gilinsky, N. L., and R. K. Bambach
1986 The evolutionary bootstrap: A new approach to the study of taxonomic diversity. *Paleobiology* 12:251–268.
- Gilinsky, N. L., S. J. Gould, and R. Z. German
1989 Asymmetries of clade shape and the direction of evolutionary time. *Science* 243:1613–1614.
- Gould, S. J.
1978 Generality and uniqueness in the history of life: An exploration with random models. *BioScience* 28:277–281.
1988 Trends as changes in variance: A new slant on progress and directionality in evolution. *Journal of Paleontology* 62:319–329.
1989 A developmental constraint in *Cerion*, with comments on the definition and interpretation of constraint in evolution. *Evolution* 43: 516–539.
1990 Speciation and sorting as the source of evolutionary trends, or 'things are seldom what they seem.' In *Evolutionary trends*, edited by K. J. McNamara, pp. 3–27. University of Arizona Press, Tucson.
- Gould, S. J., N. L. Gilinsky, and R. Z. German
1987 Asymmetry of lineages and the direction of evolutionary time. *Science* 236:1437–1441.
- Gould, S. J., D. M. Raup, J. J. Sepkoski, Jr., T. J. M. Schopf, and D. S. Simberloff
1977 The shape of evolution: A comparison of real and random clades. *Paleobiology* 3:23–40.
- Hughes, S. S.
1998 Getting to the point: Evolutionary change in prehistoric weaponry. *Journal of Archaeological Method and Theory* 5:345–408.
- Hutchinson, G. E.
1981 Random adaptation and imitation in human evolution. *American Scientist* 69:161–165.
- Jones, G. T.
1989 Comment on "Diversity, organization, and behavior in the material record" by M. J. Shott. *Current Anthropology* 30:305.
- Jones, G. T., D. K. Grayson, and C. Beck
1983 Artifact class richness and sample size in archaeological surface assemblages. In *Lulu Linear Punctated: Essays in honor of George Irving Quimby*, edited by R. C. Dunnell and

- D. K. Grayson, pp. 55-73. University of Michigan, Museum of Anthropology, Anthropological Papers No. 72.
- Kidder, A. V.
1916 The pottery of the Casas Grandes District, Chihuahua. In *Holmes anniversary volume: Anthropological essays*, edited by F. W. Hodge, pp. 253-268. Washington, DC.
- 1917 A design-sequence from New Mexico. *National Academy of Sciences, Proceedings* 3:369-370.
- Kintigh, K. W.
1984 Measuring archaeological diversity by comparison with simulated assemblages. *American Antiquity* 49:44-54.
- Kirchner, J. W., and A. Weil
1998 No fractals in fossil extinction statistics. *Nature* 395:337-338.
- Kitchell, J. A., and N. MacLeod
1988 Macroevolutionary interpretations of symmetry and synchronicity in the fossil record. *Science* 240:1190-1193.
- 1989 Response (to Gilinsky et al.). *Science* 243:1614-1615.
- Krieger, A. D.
1944 The typological concept. *American Antiquity* 9:271-288.
- Kroeber, A. L.
1940 Stimulus diffusion. *American Anthropologist* 42:1-20.
- 1944 *Peruvian archaeology in 1942*. Viking Fund Publications in Anthropology No. 4.
- Leonard, R. D., and G. T. Jones
1987 Elements of an inclusive evolutionary model for archaeology. *Journal of Anthropological Archaeology* 6:199-219.
- 1989 Editors. *Quantifying diversity in archaeology*. Cambridge University Press, Cambridge.
- Lipo, C., M. Madsen, R. C. Dunnell, and T. Hunt
1997 Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology* 16:301-333.
- Lyman, R. L.
1998 *Culture historical and biological approaches to identifying homologous traits*. Paper presented at the 63rd Annual Meeting of the Society for American Archaeology, Seattle.
- Lyman, R. L., and M. J. O'Brien
1998 The goals of evolutionary archaeology: History and explanation. *Current Anthropology* 39:615-652.
- Lyman, R. L., M. J. O'Brien, and R. C. Dunnell
1997 *The rise and fall of culture history*. Plenum, New York.
- Lyman, R. L., S. Wolverton, and M. J. O'Brien
1998 Seriation, superposition, and interdigitation: A history of Americanist graphic depictions of culture change. *American Antiquity* 63:239-261.
- McCarty, P. H., and M. F. Glass
1990 Simulation models and the interpretation of archaeological diversity. *American Antiquity* 55:521-536.
- Meltzer, D. J.
1981 A study of style and function in a class of tools. *Journal of Field Archeology* 8:313-326.
- Meltzer, D. J., R. D. Leonard, and S. K. Stratton
1992 The relationship between sample size and diversity in archaeological assemblages. *Journal of Archaeological Science* 19:375-387.
- Miller, A. I., and J. J. Sepkoski, Jr.
1988 Modeling bivalve diversification: The effect of interaction on a macroevolutionary system. *Paleobiology* 14:364-369.
- Musil, R. R.
1988 Functional efficiency and technological change: A hafting tradition model for prehistoric North America. In *Early human occupation in far western North America: The Clovis-Archaic interface*, edited by J. A. Willig, C. M. Aikens, and J. L. Fagan, pp. 373-387. Nevada State Museum, Anthropological Papers No. 21.
- Neff, H.
1992 Ceramics and evolution. *Archaeological Method and Theory* 4:141-193.
- 1993 Theory, sampling, and analytical technique in the archaeological study of prehistoric ceramics. *American Antiquity* 58:23-44.
- Neiman, F. D.
1995 Stylistic variation in evolutionary perspective: Inferences from decorative diversity and interassemblage distance in Illinois Woodland ceramic assemblages. *American Antiquity* 60:7-36.
- O'Brien, M. J., and T. D. Holland
1990 Variation, selection, and the archaeological record. *Archaeological Method and Theory* 2:31-79.
- 1992 The role of adaptation in archaeological explanation. *American Antiquity* 57:3-59.
- O'Brien, M. J., and R. L. Lyman
1998 *James A. Ford and the growth of Americanist archaeology*. University of Missouri Press, Columbia.
- 1999 *Seriation, stratigraphy, and index fossils: The backbone of archaeological dating*. Plenum, New York.

- 2000 *Applying evolutionary archaeology: A systematic approach*. Plenum, New York.
- O'Brien, M. J., R. L. Lyman, and R. D. Leonard
1998 Basic incompatibilities between evolutionary and behavioral archaeology. *American Antiquity* 63:485-498.
- O'Connell, J. F., and C. M. Inoway
1994 Surprise Valley projectile points and their chronological implications. *Journal of California and Great Basin Anthropology* 16:162-198.
- Patterson, C., and A. B. Smith
1988 Periodicity in extinction: The role of systematics. *Ecology* 70:802-811.
- Patterson, T. C.
1963 Contemporaneity and cross-dating in archaeological interpretation. *American Antiquity* 28:389-392.
- Petrie, W. M. F.
1899 Sequences in prehistoric remains. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 29:295-301.
- Petrosky, H.
1992 *The evolution of useful things*. Knopf, New York.
- Phillips, P., J. A. Ford, and J. B. Griffin
1951 Archaeological survey in the lower Mississippi Valley, 1940-1947. *Peabody Museum of American Archaeology and Ethnology, Papers* 25.
- Phillips, P., and G. R. Willey
1953 Method and theory in American archaeology: An operational basis for culture-historical integration. *American Anthropologist* 55:615-633.
- Prothero, D. R.
1985 North American mammalian diversity and Eocene-Oligocene extinctions. *Paleobiology* 11:389-405.
- Raup, D. M.
1977 Probabilistic models in evolutionary paleobiology. *American Scientist* 65:50-57.
- Raup, D. M., and S. J. Gould
1974 Stochastic simulation and evolution of morphology—Towards a numothetic paleontology. *Systematic Zoology* 23:305-322.
- Raup, D. M., S. J. Gould, T. J. M. Schopf, and D. S. Simberloff
1973 Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology* 81:525-542.
- Rensberger, J. M., and A. D. Barnosky
n.d. *On the simulation of nonrandom phyletic patterns from a random base*. Unpublished manuscript.
- Rhode, D.
1988 Measurement of archaeological diversity and the sample-size effect. *American Antiquity* 53:708-716.
- Rosenberg, M.
1994 Pattern, process, and hierarchy in the evolution of culture. *Journal of Anthropological Archaeology* 13:307-340.
- Rouse, I.
1939 Prehistory in Haiti: A study in method. *Yale University Publications in Anthropology* No. 21.
1955 On the correlation of phases of culture. *American Anthropologist* 57:713-722.
1967 Seriation in archaeology. In *American historical anthropology: Essays in honor of Leslie Spier*, edited by C. L. Riley and W. W. Taylor, pp. 153-195. Southern Illinois University Press, Carbondale.
- Rowe, J. H.
1961 Stratigraphy and seriation. *American Antiquity* 26:324-330.
- Schiffer, M. B.
1996 Some relationships between behavioral and evolutionary archaeologies. *American Antiquity* 61:643-662.
- Sepkoski, J. J., Jr., R. K. Bambach, D. M. Raup, and J. W. Valentine
1981 Phanerozoic marine diversity and the fossil record. *Nature* 293:435-437.
- Shott, M. J.
1989 Diversity, organization, and behavior in the material record. *Current Anthropology* 30:283-315.
- Signor, P. W. III
1982 Species richness in the Phanerozoic: Compensating for sampling bias. *Geology* 10:625-628.
- Spencer, C. S.
1997 Evolutionary approaches in archaeology. *Journal of Archaeological Research* 5:209-264.
- Stanley, S. M., P. W. Signor III, S. Lidgard, and A. F. Karr
1981 Natural clades differ from "random" clades: Simulations and analyses. *Paleobiology* 7:115-127.
- Stearns, S. C.
1986 Natural selection and fitness, adaptation and constraint. In *Patterns and processes in the history of life*, edited by D. M. Raup and D. Jablonski, pp. 23-44. Springer-Verlag, Berlin.
- Stein, E., and P. Lipton
1989 Where guesses come from: Evolutionary epistemology and the anomaly of guided variation. *Biology and Philosophy* 4:33-56.

- Steward, J. H.
1929 Diffusion and independent invention: A critique of logic. *American Anthropologist* 31:491-495.
- Teltser, P. A.
1995 Culture history, evolutionary theory, and frequency seriation. In *Evolutionary archaeology: Methodological issues*, edited by P. A. Teltser, pp. 51-68. University of Arizona Press, Tucson.
- Thomas, D. H.
1981 How to classify the projectile points from Monitor Valley, Nevada. *Journal of California and Great Basin Anthropology* 3:7-43.
1983a Cultural stratigraphy at Gatecliff Shelter. In *The archaeology of Monitor Valley 2: Gatecliff Shelter*, by D. H. Thomas, pp. 172-176. American Museum of Natural History, Anthropological Papers 59.
1983b Some theoretical and methodological considerations. In *The archaeology of Monitor Valley 2: Gatecliff Shelter*, by D. H. Thomas, pp. 419-440. American Museum of Natural History, Anthropological Papers 59.
- Thomas, D. H., and S. L. Bierwirth
1983 Material culture of Gatecliff Shelter: Projectile points. In *The archaeology of Monitor Valley 2: Gatecliff Shelter*, by D. H. Thomas, pp. 177-211. American Museum of Natural History, Anthropological Papers 59.
- Thompson, R. H. (Editor)
1956 An archaeological approach to the study of cultural stability. In *Seminars in archaeology: 1955*, pp. 31-57. Society for American Archaeology, Memoir No. 11.
- Uhen, M. D.
1996 An evaluation of clade-shape statistics using simulations and extinct families of mammals. *Paleobiology* 22:8-22.
- Valentine, J. W., and C. I. May
1996 Hierarchies in biology and paleontology. *Paleobiology* 22:23-33.
- Vrba, E. S.
1989 Levels of selection and sorting with special reference to the species level. *Oxford Surveys in Evolutionary Biology* 6:111-168.
- Vrba, E. S., and S. J. Gould
1986 The hierarchical expansion of sorting and selection: Sorting and selection cannot be equated. *Paleobiology* 12:217-228.
- Willey, G. R.
1945 Horizon styles and pottery traditions in Peruvian archaeology. *American Antiquity* 10:49-56.
1953 Archaeological theories and interpretation: New World. In *Anthropology today*, edited by A. L. Kroeber, pp. 361-385. University of Chicago Press, Chicago.
- Willey, G. R., and P. Phillips
1958 *Method and theory in American archaeology*. University of Chicago Press, Chicago.