Chapter 5

Classification

R. Lee Lyman and Michael J. O'Brien

INTRODUCTION

Categorization of individual phenomena has several purposes: the categories provide an information storage and retrieval system; they simplify variation into a small, manageable number of kinds more easily discussed than each individual specimen; and they provide a means of recording variation for purposes of analysis. Given these purposes, categorization occurs in virtually all endeavors. A librarian must decide if a newly published book is a work of fiction, a work of history, or a work of historical fiction. Astronomers must decide if a newly discovered celestial body is a star, a planet, a moon, an asteroid, or something else entirely. Pedologists must decide which category of soil occurs in a particular area. When you buy a new car, you must decide if you want to drive a Ford, a Chevrolet, or a Toyota.

Anthropologists have used a host of sorting systems to simplify, organize, and analyze the materials they study. Nineteenth-century philologists categorized languages in a manner still used to assess the evolutionary development and relations of modern languages. Late in the nineteenth century, Americanist anthropologists applied the notion of culture areas to sort collections of artifacts for museum displays. Nineteenth-century anthropometrists sorted people in various ways, one of the better-known ones using the length–width ratio of the skull—the "cephalic index" developed by Anders Retzius in 1842. A person is brachycephalic (short, broad head) if his skull is 82% as wide as it is long, dolichocephalic (long, narrow head) if his skull is 77% as wide as it is long, and mesocephalic if the length–width ratio falls between those values. These categories were sometimes used by archaeologists between 1900 and 1950 to distinguish among prehistoric groups of people. As we enter the third millennium, many of these sorting schemes have been discarded as unrealistic, invalid, or unusable.

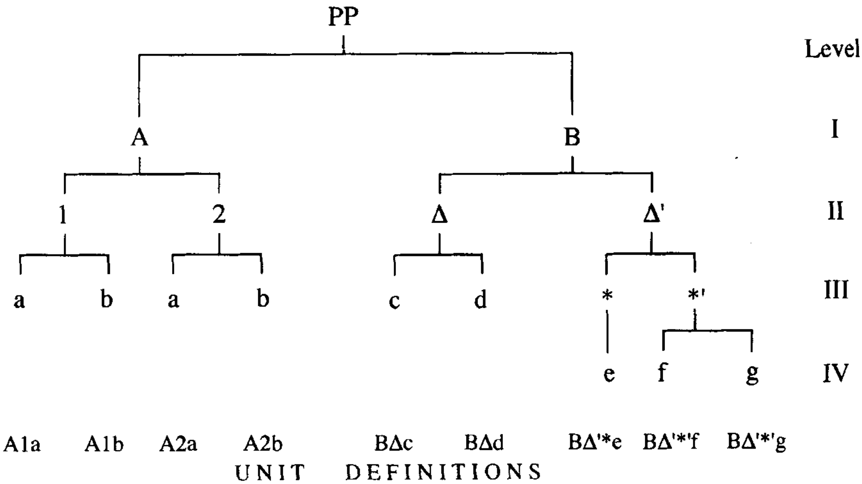
Given the ubiquity of categorization in everyday life, it might be surprising that it took many years, often many decades, to develop the categorization systems used in various everyday, scientific, and humanistic endeavors. The categories earth, air, fire, and water were once adequate for sorting phenomena by material type, but these were replaced when the first periodic table was produced by Russian chemist Dmitri Mendeleev in 1869. He arranged elements by increasing atomic weight, whereas the modern periodic table arranges elements by increasing atomic number. The original system of categorizing soils—phenomena more familiar to archaeologists—went through numerous modifications and revisions as pedologists attempted to develop a system that served some useful purpose. Similarly, the system for categorizing the strata of the geological record has undergone regular revision and has on occasion been modified for archaeological purposes (Gasche and Tunca 1983). Scientific categorizations are always susceptible, and in fact should be amenable, to change.

The systems, processes, and results of categorization are referred to by a plethora of terms that are not always synonymous, even within any given discipline, because they typically have no commonly agreed-on meaning. Standard dictionaries do not always help clarify things because their included definitions may present ideals rather than the meaning of the terms as they are used in particular real-world situations. *Webster's Seventh New Collegiate Dictionary* (1967) defines "classification" not only as "the act or process of classifying," where classifying comprises "assigning to a category," but also as "a systematic arrangement in groups or categories according to established criteria; *specif*: taxonomy." Note that the process of the first definition demands the categories of the second definition. This same dictionary defines "taxonomy" as "the study of the general principles of scientific classification: systematics"; "systematics" in turn is defined as "of, relating to, or concerned with classification; *specif*: taxonomic."

We define **classification** as the creation of new units and the modification and revision of old units by stipulating the necessary and sufficient conditions for membership within a unit (Dunnell 1971). The term **unit** denotes a conceptual entity that serves as a standard of measurement (Ramenofsky and Steffen 1998). A centimeter is a unit constructed explicitly to measure linear distance; the degrees on a compass are units constructed explicitly to measure geographic direction or orientation. As conceptual entities, units must be explicitly defined if they are to be usable for measuring (characterizing, describing, classifying) phenomena. Units can be specified at any scale. Phenomena to be classified may comprise discrete objects such as projectile points or organisms; they may comprise attributes of discrete objects such as the bits of temper in ceramics or the genes in organisms; or they may comprise sets of discrete objects such as aggregates of tools variously termed assemblages or tool kits, or populations of organisms variously termed faunas, floras, or communities. Phenomena are classified on the basis of their **form**, using properties such as size, shape, color, frequency, and material. We use the term "form" and its derivatives throughout to denote **attributes**, or **characters**, of phenomena. Attributes used to classify phenomena occur at a finer, less inclusive scale than the phenomena themselves; attributes are formal properties of the phenomena being classified.

Systematics is the study of diversity of the phenomena of interest, irrespective of the scale or kind of phenomena, and the sorting of that diversity into sets such that like goes with like. The goal of systematics in all disciplines, including anthropology and archaeology, is to sort phenomena into sets of individuals that are in some sense similar; each set should be internally homogeneous such that within-group variation is analytically meaningless and between-group variation meaningful, where a **group** is an empirical unit comprising one or more specimens. Similar phenomena are conceived of as being not only formally similar but as being similar in other ways as well. **Affinity** refers to the relation between formally similar specimens within a group or between groups of formally dissimilar specimens (Simpson 1945). The relation specified when one states that every specimen of kind A has an affinity with every other specimen of that kind, and that specimens of kind A have a different affinity with those of kind B, often is of a particular sort. Multiple kinds of things may be affines because they are close in time, in function or purpose, in symbolism, in ancestry, or in terms of several of these or something else. Specifying and measuring a particular kind of affinity is the ultimate goal of classification, irrespective of discipline.

Typology and **type** are most often found in the archaeological literature, where the former is used as a synonym for systematics and classification and the latter as a synonym for unit (Dunnell 1986). **Taxonomy** concerns the study of theories of classification, their bases, principles, procedures, and rules (Simpson 1961). A **taxonomic classification** is a hierarchical arrangement in which characters are weighted and considered in order of their suspected importance such that units at one rank include parts that are units at lower ranks (Valentine and May 1996). Figure 5.1 shows an example of a four-level taxonomic classification. The particular weighting of attributes reflected by the order in which they are considered influences the nature of the resulting units, and thus a taxonomic classification can be difficult to use (Allen 1996). A **key** is a set of particular attributes arranged such that individual specimens may be identified as belonging to one category or another. A **paradigmatic classification** is a multidimensional arrangement in which each dimension comprises a particular category of attribute (e.g., color, length, material) and its various states (e.g.,



PP = projectile point

A = notches present; B = notches absent, stemmed

1 = side notch; 2 = basal notch

a = shallow notch; b = deep notch

Δ = constricted stem; Δ' = straight stem

c = curved based; d = pointed base

* = short stem; *' = long stem

e = straight base; f = straight base; g = convex base

Figure 5.1. A hypothetical four-level taxonomic classification for projectile points. The four units on the left and the two in the center are each defined by three attributes; the three units on the right are each defined by four attributes. The exclusion of a fourth definitive attribute for the six units on the left indicates either that base shape does not vary or is analytically insignificant, but it is unclear which applies or if both apply. Identifying specimens as members of a particular unit must consider attribute levels I–IV in that order, as reversing the order of levels I and IV would significantly alter unit definitions.

for length, 0.1–1.0 cm; 1.1–2.0 cm; 2.1–3.0 cm; etc.). No dimension or attribute state is given more weight than any other, and all dimensions play a role in the identification of a specimen as belonging to a particular unit (Dunnell 1971). Figure 5.2 shows a hypothetical, two-dimensional paradigmatic classification of projectile points, where dimension 1 is the location of notches and dimension 2 is base shape. All dimensions and attributes are considered of equal weight and thus the order in which they are considered does not influence the nature of the units.

Dimension 1—Notch Location

Dimension 2—Base Shape	1. Side	2. Corner	3. Basal	4. None
A. Convex	1A	2A	3A	4A
B. Concave	1B	2B	3B	4B
C. Straight	1C	2C	3C	4C

Figure 5.2. A hypothetical two-dimensional paradigmatic classification for projectile points. Dimension 1, base shape, has three attribute states; dimension 2, location of notching, has four attribute states. Twelve units result from intersection of the dimensions (three-by-four); definitions of each unit are shown in each cell. In contrast to taxonomic classification (Figure 5.1), no dimension or attribute is weighted as more or less important than any other, and specimens are readily identified as members of a particular unit because there is no requisite order to consideration of dimensions and attributes.

To be useful, a classification must allow one to do some analytical work. One implication is that traditional archaeological types such as "Clovis points" and "Dalton points" may not satisfactorily perform the analytical work we ask of them today. A second implication is that a set of phenomena can be classified in a virtually infinite number of ways, although we are aware of very few examples of a collection of artifacts being classified in more than one way even when the collection is used to answer disparate analytical questions. A third implication is that the analytical validity of the units produced by classification must be testable. Do they measure the kind of affinity sought? Recent innovative work has involved testing the reliability of identifying specimens as members of a particular unit (Whittaker et al. 1998). Increased technological sophistication in the laboratory has produced insights into attributes of artifacts of finer scale and greater resolution and sometimes of previously unknown attributes. These important advances beg two questions. Of what utility is a classification that is reliably applied from classifier to classifier if it fails to measure the kind(s) of affinity within and between units that the analyst seeks? And, of what use is the fact that we can measure something to the nearest tenth of a millimeter rather than to the nearest millimeter if such instances of finer resolution are not geared toward detecting variation that is of analytical importance?

These questions highlight a final implication of the fact that a classification must allow one to do some analytical work. How do we know in the first place which attributes we should measure, and how do we know we should be measuring, say, a variable to the nearest tenth of a millimeter, the presumption being that measurement to the nearest millimeter is insufficient? There must be some **theory**—a set of things and statements about how those things interact that provide explanations—that guides analysis, because it is theory and its derivative propositions that suggest which attributes are relevant and at what scale they should be measured.

HISTORY

In the fourth century B.C., Aristotle—himself classified as "the first great classifier" (Mayr 1968: 595)—sought to classify biological organisms along a single graded scale known as the *scala naturae*, or "Great Chain of Being," according to their degree of perfection. To accomplish this, Aristotle sought the underlying essence—the essential characteristics—of each kind of organism. In the twentieth century his metaphysic came to be known as **essentialism**, or **typological thinking** (Mayr 1959). This ontology heavily influenced all classifications until Charles Darwin proposed an alternative, today termed **materialism**, or **population thinking**. Darwin focused on the uniqueness of phenomena and thus, although the basic form of individuals within a set of similar phenomena can be captured by, say, a statistical average, such measures of central tendency are abstractions and in no sense real. Alternatively, typological thinking holds that types are real and fixed—a statistical average comprises an essence—and variability between individuals within a kind has no analytical importance.

Either the ontology of essentialism or that of materialism underpins all classifications. This leads to a misunderstanding of the meaning of particular classifications, but it does not mean that one ontology is always preferred over the other. Essentialism is advantageous when prediction and laws are desired about how kinds of things interact. The things and their interactions will always, regardless of their positions in time and space, be the same because the essential properties of the things are the same. The periodic table of chemistry is founded in essentialism. When history is the focus of study, materialism is preferred because although international conflicts, plagues, droughts, and other historical events recur, each particular event is unique in potentially critical attributes, despite the fact that we can construct a classification of them.

The definition of **evolution** as change in the frequencies of phylogenetically related variants demands a materialist ontology because the **processes** (actions that produce a result) of Darwinian evolution—transmission, replication, drift, and natural selection—concern variants. Transmission of information, whether genetic or conceptual, as in teaching an individual how to decorate ceramics, is what results in **heritable continuity** because attributes are replicated. **Drift** comprises differential replication as a result of transmission error or lack of fidelity in **inheritance** (the movement of information—cultural or genetic—from one organism to another), and **natural selection** comprises differential replication of more-adapted and less-adapted forms.

Within paleobiological classification we find problems parallel to those in archaeology and also some possible solutions. Biology and sociocultural anthropology are sister disciplines; both study their subject phenomena when those phenomena are operating. Paleobiology and archaeology are also sister disciplines; both study the prehistory and evolutionary development of organisms, and both grew out of similar disciplinary antecedents (O'Brien and Lyman 2000). All four disciplines have some common goals—to explain the diversity of phenomena of interest and to write and explain a history of the development of that diversity in evolutionary terms. But biologists study organisms and species, and anthropologists study people and **culture** (socially transmitted behavior) or cultures; paleobiologists study fossils, and archaeologists study artifacts. Differences in the materials studied and the desire of paleobiologists and archaeologists to emulate biologists and anthropologists, respectively, is where problems in classification originate.

Biological Systematics

During the middle years of the twentieth century, biologists regularly lamented that systematics, given its central role in biological inquiry (e.g., Huxley 1940; Mayr 1968; Simpson 1945), had not received the recognition that it should have. These laments marked a shift in the focus of biological systematics after the 1940s neo-Darwinian Synthesis from the notion of a species as a morphological unit—what Simpson (1945: 3) referred to as "archetypal" classification—to the biological concept of a **species** as a reproductively isolated population of interbreeding organisms (Mayr 1942). Despite the fact that there are nearly two dozen distinct species concepts presently under discussion (Mayden 1997), many biologists and philosophers of biology recognize that a species is a unit constructed for some analytical or applied biological purpose (Hull 1997; Mayr 1968). Those purposes might be for managing biodiversity or for studying the phylogenetic history of a group of organisms. One's analytical goal dictates which one of the several available species concepts, and thus which set of units, is the most appropriate. Given an interest in evolutionary history, some argue that conceiving of a species as a reproductively isolated set of organisms fulfills the requirement of being an evolutionary unit. What did paleobiologists do in light of the new biological conception of species?

In the nineteenth and early twentieth centuries, paleontologists were much like their contemporaries, whom modern archaeologists term "antiquarians," in terms of their archetypal classifications. Subsequent to the neo-Darwinian Synthesis, initial efforts were made to rewrite Darwinian evolutionary theory so that it would explain the data derived from the paleontological record (e.g., Simpson 1943, 1944; see Eldredge [1985, 1989] for historical overviews). These efforts eventually resulted in the initiation of a new journal in 1975—*Paleobiology*—devoted to studying and explaining the fossil record in Darwinian terms and to rewriting, fine-tuning, and expanding that theory in terms of paleontological data (e.g., Eldredge 1999). To reach this point required a concomitant change in paleontological systematics and in the language of evolutionary theory. And here is where archaeologists can learn a valuable lesson.

In the 1940s and 1950s paleontologists worried about how they were going to operationalize and thus incorporate the biological-species concept into paleontology (Sylvester-Bradley 1956). Fossils comprising a population of organisms, after all, did not variously interbreed. Yet paleontologists wanted to study species because, based on the biological-species concept, they were thought to have biological meaning and thus to be the units of evolution. Further, **lineages**—evolutionary continua or lines of heritable continuity—had to be divided more or less arbitrarily into chunks. Those chunks were termed either "chronospecies," signifying that the temporal boundaries of the chunks were arbitrary, or "morphospecies," signifying that the units were arbitrarily delimited formal ones that may not comprise a reproductively isolated set of organisms. This awkward state of affairs resulted because paleontologists were attempting to rewrite the paleontological record in biological terms. In words more familiar to archaeologists, they were trying to reconstruct the static paleontological record into a dynamic biological system.

These difficulties were resolved, not without debate, after traditional evolutionary theory was rewritten in paleontological terms and the biological species concept was rendered applicable to the paleontological record. The rewriting comprised the punctuated-equilibrium version of Darwinian evolutionary theory (Eldredge and Gould 1972; Gould and Eldredge 1977, 1986), produced by paleontologists who viewed (1) fossil-species units as equivalent to extant biological species because of the formal stasis of each and (2) formal variation in fossils as a result of genetic variation (Eldredge 1999). Traditional evolutionary theory, written by biologists, viewed evolution as a seamless, continuous, gradual process necessitating "arbitrary" delimitation of fossil species. But formal stasis over long time periods was empirically evident to some paleontologists. Granting that these static units were equivalent to extant biological species, the implications were significant for evolutionary theory. Hence, that theory and its attendant units were rewritten in paleontological terms, which is not to say that biological concepts were discarded (Gould and Eldredge [1993] and references therein). Indeed, they form a major part of paleobiology.

Archaeological Systematics

Hallmark events in archaeological systematics include the debate between James Ford (1954a, 1954b) and Albert Spaulding (1953, 1954) in the 1950s (see O'Brien and Lyman [1998] for detailed discussion) and the debate between Lewis and Sally Binford (Binford and Binford 1966) and Francois Bordes (1961) in the 1960s. Both debates focus in part on what kind of affinity an archaeological "type" signifies. The term "type" in archaeology is in many ways parallel to the term "species" in biology and paleobiology. Archaeologists have long desired types that not only allow analytical work to be performed but also are culturally meaningful. That is, they want types (1) that monitor adaptive or functional variation; (2) that serve as index fossils for purposes of stratigraphic correlation; and (3) that signify a particular ethnic, linguistic, or cultural group or some form of social organization or political structure.

During the last third of the nineteenth century and the first fifteen or so years of the twentieth century, anthropologists and archaeologists sought classifications that were universally applicable and which resulted in specimens being placed in their "proper" types. Many of these units can loosely be categorized as functional—weapons of war, items of adornment—but they were informed by common sense and typically only by accident had any useful archaeological meaning. Discussions of the hows and whys of classification were noticeably rare (Dunnell 1986). This situation changed somewhat after 1910 when it was discovered that if types were constructed in particular ways, they had a particular kind of distribution in time and space. The pursuit of what were later termed "styles," or "historical types," became the focus of classification efforts and resulted in and underpinned the emergence of what came to be known as the culture-history paradigm (Lyman and O'Brien 1999; Lyman et al. 1997, 1998).

From the 1930s into the 1960s, archaeological types that allowed analytical work to be performed were warranted by Americanist archaeologists with ethnographic observations, which resulted in such explanatory axioms as the popularity principle that underpins **frequency seriation** (Lyman et al. 1997; frequency seriation involves arranging multiple assemblages of artifacts based on the similarities of the relative frequencies of the included types). Derivation of these kinds of **explanations** (reason-giving statements) became axiomatized with explicit use of ethnographic **analogy** beginning in the 1950s and 1960s (e.g., Ascher 1961; Binford 1967). Simply put, analogy comprises the reasoning that if two phenomena visibly share some attributes, then they share other attributes as well; typically the latter attributes are invisible in the archaeological specimen but visible in the ethnographic analog. This is thought to provide types with the desired anthropological, cultural, and human-behavioral significance; the reasoning is that if two units share some visible formal characters, then they share other characters that generally are visible in the modern analog and invisible in the archaeological material. However, the modern analog, whether at the scale of a discrete object or at the scale of sociopolitical organization, typically comprises an empirical generalization founded on a sample of ethnographic observations rendered as a type rather than on explicit specification of the necessary and sufficient conditions for membership in a unit derived from theory. Thus, we have units such as "tribes" and "foragers." In marked contrast, the biological-species concept was constructed on the basis of evolutionary theory; the concept was applicable to the paleontological record only after the theory was reworded in terms relevant to the fossil record.

The historical types constructed by A. L. Kroeber, Nels Nelson, A. V. Kidder, and Leslie Spier in the second decade of the twentieth century were of a kind that allowed them to measure time and were founded on a materialist ontology (Lyman et al. 1997; O'Brien and Lyman 1999). These types had to pass the historical-significance test (Krieger 1944). That is, they had to occur during only one span of time (length of duration was unspecified) and in a geographically limited area, the latter to ensure that temporal variation in form was being measured rather than geographic variation in form (Lyman and O'Brien 2000; O'Brien and Lyman 1999). In short, the analytical utility of historical types had to be tested—they had to have demonstrable temporal affinity. That the types were analytical units built by trial and error without the benefit of theory escaped notice. As a result, that some types occasionally overlapped in time—occurred in multiple assemblages—and thus signified heritable continuity (because unit similarity was **homologous**, or the result of shared ancestry) was noted, and such phenomena soon became known as traditions. This was largely ignored, however, in favor of studying temporally discrete and discontinuous units termed phases, cultures, and the like.

Most practitioners agreed that the analytical utility of the types for measuring time had to be tested, but the test was narrowly focused on whether types measured the passage of time. If they did, it was inferred—in the absence of theory—that they also measured the heritable continuity demanded by an evolutionary lineage. One could speak of the "historical relatedness" of archaeological units (Willey 1953), by which was meant that if two or more units were formally, temporally, and geographically similar, they were phylogenetically related. One kind of affinity—evolutionary—was inferred from three others—formal, temporal, and spatial. In the absence of independent chronological data, phylogenetic affinity was the implicit assumption underpinning the temporal ordering of artifacts, imparting a degree of circularity to the reasoning. This is why stratigraphic excavation quickly became favored over chronological techniques such as seriation (Lyman and O'Brien 1999).

The early twentieth-century explanation given by archaeologists for historical types having the spatio-temporal distributions they did was that the types reflected the passage of time because they represented ethnographically observable changes in taste, fashion, or popularity. Therefore, it was suspected that those types had emic meaning and also some sort of empirical reality—they were more or less accurate reflections of past peoples' mental templates. The validity of such emic units, however, could not be tested empirically. Part of the problem resided in the often murky procedure by which historical types were constructed. Procedural murkiness was attacked explicitly by Spaulding (1953), who advocated statistical tests of the significance of attribute combinations. But such tests only show statistically significant attribute combinations and rest on an essentialist ontology. In the absence of units constructed on the basis of explanatory theory, what those attribute combinations signify—the kind of affinity they measure—is a matter of inference.

CONTEMPORARY USES

Despite the importance of classification to any scientific undertaking and the desire of archaeologists to be scientific, the archaeological literature contains few detailed studies of theories of classification, their bases, principles, procedures, and rules. This is so for two reasons. First, the analytical goals of archaeology today are virtually as diverse as the genotypes of professional archaeologists. Second, Americanist archaeologists, at least, are trained as anthropologists and are steeped in ethnological and cultural theory. In part because of these two facts, archaeologists use the theories they know—anthropological theories—to explain the archaeological record. Because that record is not anthropological it must somehow be made anthropological—meaning it must be turned into a *cultural* record. Further, if archaeological units are to be explained with anthropological theory, they must have anthropological meaning and relevance (e.g., Phillips and Willey 1953). This is precisely the problem that paleontologists struggled with in the middle decades of the twentieth century (1940–1970) as they tried to use the neo-Darwinian version of evolutionary theory written in biological terms such as the biological-species concept. Paleontologists eventually found a way out of this difficulty by rewriting the theory in paleontological terms; some archaeologists are attempting to implement this kind of solution (O'Brien and Lyman 2000), whereas others continue to try to make the original solution work (Spencer [1997] and references therein).

We believe that a version of Darwinian evolutionary theory rewritten in archaeological terms would resolve many of the problems archaeologists face today as well as provide theoretically informed answers to many of their questions (Lyman and O'Brien 1998). Just as the paleontologically applicable version of evolutionary theory known as punctuated equilibrium incorporates elements of the biological version of that theory, so too must an archaeological version of evolutionary theory incorporate various elements of anthropological theory. Concepts such as artifact and processes such as diffusion are examples; the former are treated as parts of the human phenotype and the latter as transmission (the movement of information from one organism to another). Because theory demands classifications of various sorts, part of that rewriting of the theory must attend taxonomy. The methods of classification must be explicitly clear, and why certain methods were used must be equally clear. We prefer paradigmatic classification for reasons mentioned above and elaborated elsewhere (O'Brien and Lyman 2000).

Because theory is the source of our ideas on causes of affinity, it has to be the final arbiter of which units are applicable for which kinds of analytical jobs. Theory dictates which variables out of the almost infinite number that could be selected are actually chosen by the analyst for measurement, and it may specify the values (attributes) those variables should take in our classifications. The variables and values chosen are the units used to construct types, and as such they are conceptual, or **ideational, units** (Dunnell 1971, 1986). The specimens we classify are **empirical**, or **phenomenological, units.** In our view, one major problem in archaeological systematics is the confusion between ideational and empirical units. Perhaps this is because ideational units—**classes**—can be **descriptional units,** used merely to characterize or describe a property or a thing, or they can be **theoretical units,** which are created for specific analytical purposes. For example, in light of a proposed causal relation between function and edge angles of stone tools, edge-angle units such as 1–30°, 31–60°, and 61–90° could be constructed as theoretical units. A theoretical unit is an ideational unit that has explanatory significance specifically because of its theoretical relevance.

Conflation of ideational and empirical units leads to the erroneous conclusion that types have emic-like significance. This conclusion also results from how ideational units are defined. An **intensional definition** comprises the necessary and sufficient conditions for membership in a unit; it explicitly lists the definitive features that a phenomenon must display to be identified as a member of the unit. The significant characteristics of the unit are derived from theory; there is no necessary reference to real, empirical specimens when the unit is constructed other than to specify that, say, projectile points rather than ceramic sherds comprise the phenomena to be classified. Thus the three classes of edge angle mentioned above—1–30°, 31–60°, and 61–90°—derive from our understanding of the mechanics of (stone) tools;

mechanics indicate some edge angles are necessary for efficient performance of some functions, whereas other edge angles are necessary for other functions.

An **extensional definition** also comprises the necessary and sufficient conditions for membership in a unit, but it is derived by enumerating selected attributes shared by the unit's members. That is, the definition is based on observed attributes of the existing members of a unit. The significant characteristics of extensionally defined units are not theoretically informed in an explicit manner because the group of specimens was formed by some murky process prior to the specification of the (extensional) definition. Most types traditionally employed in archaeology are extensionally defined units formed when an analyst subdivides a collection of artifacts into smaller piles based on perceived similarities and differences. (The procedure is murky because what one analyst chooses to perceive may be different from the choices of another analyst.) A summary of the central tendencies of the members of each pile, or a statement on the normal appearance of specimens in each pile, comprises the definitive criteria of a type. Unit definitions depend entirely on the specimens examined. Thus, we cannot know if such extensionally defined units are comparable in terms of the kind of affinity we hope they measure.

An example of the troublesome results of this procedure concerns the early history of the systematics of Paleoindian-period projectile points in the American Southwest. What we would now term "Clovis" points were first referred to by the term "Folsomoid" and the like (LeTourneau 1998). The problem was, there were no definitive criteria for distinguishing between Clovis and Folsom points until sufficient specimens had been examined to detect, by trial and error, which attributes allowed their consistent discrimination in form (and time).

The goals of evolutionary archaeology comprise writing the histories of cultural lineages and explaining why those histories have the forms they do (Lyman and O'Brien 1998). To do this, we need to measure two distinct kinds of affinity, and thus our units need to be of two kinds. We need units that monitor heritable continuity, or what have been termed **styles** or less commonly, **stylistic units**. This kind of unit ensures that we are documenting lineages rather than merely temporal sequences, given that stylistic similarity is by definition the result of transmission (Lipo et al. 1997). Change can occur within a lineage as the result of a lack of perfect fidelity in replication, which explains why various seriation techniques and percentage stratigraphy work as chronometers (Lyman and O'Brien 2000; O'Brien and Lyman 1999, 2000). Such units can be used to track the transmission pathways requisite to answering evolutionary questions, but they also allow the identification of prehistoric interaction over geographic space (Lipo et al. 1997). These units allow identification of prehistoric transmission and thus the writing of phylogenetic history. The analytical utility of styles was recognized early on in anthropology by E. A. Hooton, who noted that "non-adaptive bodily characters" are those that "have been derived from their common descent" and that only these characters are the "result of the same ancestry" (Hooton 1926: 76, 77). Hooton was accounting for the fact that adaptive characters may result from convergence—the derivation of similar solutions to a problem—rather than from shared ancestry.

Evolutionary archaeology also requires units that measure functional variation in artifacts, because this kind of unit allows us to call on natural selection as another process of change (e.g., O'Brien et al. 1994). This is not to say that functional units will not measure transmission; like stylistic units, **functional units** can produce the lenticular curves expected of frequency seriation. However, the transmission and replication of functional units is mediated by selection, often to such an extent that lenticular frequency distributions do not result. Further, because the transmission of functional units is mediated by selection, they are expected to have rather different distributions in time and space than stylistic units (see below). Finally, adaptive convergence may produce units that are similar, and together these may have spatio-temporal distributions that resemble those produced by heritable continuity. Functional properties are those that influence the efficiency with which a task is performed; a projectile point must be pointed to penetrate a prey animal, and it must also be sharp enough to cut sufficient tissue that the animal bleeds to death. Attributes of hafting may also be functional, and although they influence the pierce–cut function, they appear to relate more to the weapon-delivery system (Hughes 1998).

CASE STUDIES

Because they measure different evolutionary processes, both stylistic and functional units are required of any attempt to explain the archaeological record in Darwinian terms. This means that traditional archaeological types often do not comprise units appropriate to the questions evolutionary archaeologists ask. This plus the fact that the theory is still being rewritten in archaeological terms means that a great deal of basic work must go into producing a substantive result. Such results are beginning to appear with some regularity (e.g., Hughes 1998; the papers introduced by Kornbacher and Madsen [1999]; and references therein). Here, we summarize a study that illustrates the innovative insights that can be gained using evolutionary theory and traditional archaeological units. We also describe a study that indicates how a general category of artifact might be classified more than one way, and the different kinds of results that can be produced from each when the classifications are constructed on the basis of evolutionary theory.

Tracking an evolutionary lineage—writing evolutionary history—was the goal of the culture historians (Lyman et al. 1997). Contributing to that end, frequency seriation was used as a chronological tool. This technique works as a chronological tool precisely because it monitors heritable continuity (O'Brien and Lyman 1999), and thus it can be used for purposes other than chronology building. Transmission has both a temporal and a spatial aspect. That is, it takes place over time, and it involves a sender and a receiver who occupy different spatial loci and who may be able to move after sending or receiving information. Frequency seriation requires that the included historical types are relatively limited in their spatial distribution but are relatively less limited in their temporal distribution (O'Brien and Lyman [1999] and references therein). Carl Lipo and his colleagues (1997) explored this aspect of historical types with computer simulations, explicitly basing their simulations on notions of heritable continuity and lineages of artifacts. The simulations matched expectations drawn from evolutionary theory that adaptively neutral units—styles—would produce the familiar battleship-shaped frequency distributions not only over time but within limited spatial units as well.

Lipo et al. (1997) used a set of historical types that had been constructed in the 1940s to seriate collections from numerous sites in an area of the southeastern United States approximately 75 kilometers by 140 kilometers. Because the 50-year-old seriations were successful and later confirmed by independent chronological evidence, this suggested that the types were in fact styles; that is, they were adaptively neutral, or nonfunctional. Lipo et al. used frequency seriation to group collections based on their similarities in terms of the relative abundances of included types, and identified sets of sites that seemed to have been occupied by human groups that interacted with one another. Although such a result may seem trivial in some respects, it most definitely is not for several reasons. First, theory dictated which kinds of units—styles—should be used. Second, archaeology, not anthropology, provided an analytical technique—frequency seriation—that was implicitly yet strongly founded in evolutionary theory. Third, the simulations were based on evolutionary theory and indicated the sort of archaeologically visible signature that would result from interaction of—transmission between—people occupying different positions on the landscape.

Previous efforts to measure such interaction rested on concepts such as horizon styles, and so depended solely on how items were classified. If two artifacts from spatially distinct sites were categorized as members of the same horizon style, then the implication was that people at one site had "influenced" people at the other site (Phillips and Willey 1953). The inference was based on common sense and founded in ethnographic observations of diffusion rather than in theory, and there was no way to test such inferences. Lipo et al.'s (1997) contribution comprises a major step toward rewriting evolutionary theory in archaeological terms while simultaneously incorporating what is known ethnologically also in archaeological terms, particularly with respect to artifact units.

Earlier we indicated that particular sets of artifacts are seldom classified in more than one way. We also indicated that to be successful, evolutionary archaeology required two kinds of units—what we termed styles and functional variants. Melinda Allen (1996) classified a set of fish hooks from East Polynesian sites in terms of what she argued were stylistic attributes, and she also classified the same set of hooks in terms of what she argued were functional attributes. Stylistic attributes are those found on the proximal end—the "head"—of the hook where the fishing line is attached. In general, functional attributes are those influencing hook performance when the line is pulled and involve the curvature of the hook. Specifically, they are the relation between the alignment of the hook shank and the hook point. Based on the general evolutionary model of how adaptively neutral traits and adaptively significant (functional) traits should alter over time, Allen (1996) argued that the relative frequencies of stylistic features captured by intensionally defined classes should fluctuate gradually and unimodally through time, whereas functional classes should vary predictably with environment. The classes she defined do so. Further, the stylistic classes vary independently of environmental setting and independently of functional features, lending support to the notion that these classes are in fact stylistic. In contrast, the functional classes do not vary consistently in their frequencies over time.

Like Lipo et al.'s (1997) study of pottery, Allen's (1996) study of fish hooks may seem trivial. But also like Lipo et al.'s study, it most decidedly is not. The attributes used for both classifications were selected on the basis of expectations derived from the theory of evolutionary descent with modification written in archaeological terms. These expectations were tested by assessing if the hypothesized stylistic classes and the hypothesized functional classes behaved—had temporo-spatial distributions—the way the theory suggests each kind of unit should. Finding that they did behave as they should allowed Allen to then detect evidence of interaction between peoples occupying different islands using the stylistic classes, and to measure adaptive change using the functional classes. Allen did not attempt to reconstitute the archaeological record into something an ethnographer would recognize and then explain that record in anthropological terms. Rather, she used Darwinian evolutionary theory to inform the nature of her classification units and to aid her explanations of the archaeological record in terms an evolutionist, an archaeologist, and an anthropologist would recognize.

FUTURE IMPORTANCE

In 1964, Supreme Court Justice Potter Stewart made the following statement: "I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description, and perhaps

I could never succeed in intelligibly doing so. But I know it when I see it." The "it" Justice Stewart was referring to was hard-core pornography, and his statement underscores the importance of classification. Would he categorize various of the top money-earning movies released in the United States in the 1990s as "hard-core pornography"? Neither he nor we can answer this question without an explicit definition of what hard-core pornography is because the answer resides in the classification system one employs. Today, pornography is defined by "community standards," resulting in as many definitions as there are communities. Such will simply not do in archaeology, where we must agree on when a particular artifact is a Clovis point or a Dalton point and when a particular ceramic sherd is a Barton Incised or Ranch Incised.

The status quo in archaeological systematics is largely being maintained because of a general failure of archaeologists to recognize the significance of classification to archaeological endeavors. Many of the historical types constructed and tested decades ago by culture historians are still referred to by name, and they are still used for myriad analytical purposes, despite the fact that they may be inappropriate for such purposes. Most of these types were built specifically to measure the passage of time, a purpose that many of them serve quite well. To expect them to be equally capable of doing other sorts of analytical work is ill-advised, particularly in the absence of tests showing that they are in fact capable of such. There are interesting parallels between paleobiological systematics and archaeological systematics, but the literature on the former is less well-known to archaeologists than the literature on the latter. And yet there are important lessons to be gleaned from paleobiology.

First, and we think most important, archaeologists must have an analytical goal. This will suggest the kinds of units to construct and the requisite scales of resolution. Second, the analytical validity and utility of the units must be testable—that is, it must be empirically ascertained if the units do what they are supposed to do. This leaves open the question of analytical goal. One traditional goal of anthropology and archaeology has been to document and explain the diversity of cultural manifestations by phrasing questions and hypotheses in historical terms. That goal still seems to be a legitimate one, as does the basic approach to it, although we favor casting the questions, hypotheses, and possible answers in Darwinian rather than strictly sociocultural terms (Lyman and O'Brien 1997, 1998; O'Brien et al. 1998).

Archaeologists would be well-advised to heed the words of philosopher David Hull (1970: 32):

> The two processes of constructing classifications and of discovering scientific laws and formulating scientific theories must be carried on together. Neither can outstrip the other very far without engendering mutually injurious effects. The idea that an

extensive and elaborate classification can be considered in isolation from all scientific theories and then transformed only later into a theoretically significant classification is purely illusory.

The object lesson of the species problem in paleobiology underscores Hull's central point: systematics is more than fundamental to scientific inquiry; it is critical. To ignore it, to designate its status as second class, and to use it without thinking, all endanger the success of scientific research.

ACKNOWLEDGMENTS

John Darwent and Steve Wolverton provided useful discussion and comments on earlier versions of this chapter. E. J. O'Brien, J. Hart, J. Terrell, and an anonymous reviewer made us aware of various problems in a first draft.

REFERENCES

Allen, M. S. (1996). Style and function in east Polynesian fish-hooks. *Antiquity* 70: 97–116.

Ascher, R. (1961). Analogy in archaeological interpretation. *Southwestern Journal of Anthropology* 17: 317–325.

Binford, L. R. (1967). Smudge pits and hide smoking: The use of analogy in archaeological reasoning. *American Antiquity* 32: 1–12.

Binford, L. R., and S. R. Binford. (1966). A preliminary analysis of functional variability in the Mousterian of Levallois Facies. *American Anthropologist* 68: 238–295.

Bordes, F. (1961). Mousterian cultures in France. *Science* 134: 803–810.

Dunnell, R. C. (1971). *Systematics in Prehistory*. New York: Free Press.

Dunnell, R. C. (1986). Methodological issues in Americanist artifact classification. In M. B. Schiffer (ed.), *Advances in Archaeological Method and Theory*, vol. 9. Orlando, FL: Academic Press, pp. 149–207.

Eldredge, N. (1985). *Unfinished Synthesis: Biological Hierarchies and Modern Evolutionary Thought*. Oxford: Oxford University Press.

Eldredge, N. (1989). *Time Frames: The Evolution of Punctuated Equilibria*. Princeton, NJ: Princeton University Press.

Eldredge, N. (1999). *The Pattern of Evolution*. New York: Freeman.

Eldredge, N., and S. J. Gould. (1972). Punctuated equilibria: An alternative to phyletic gradualism. In T.J.M. Schopf (ed.), *Models in Paleobiology*. San Francisco: Freeman, Cooper, pp. 82–115.

Ford, J. A. (1954a). Comment on A. C. Spaulding's "Statistical techniques for the discovery of artifact types." *American Antiquity* 19: 390–391.

Ford, J. A. (1954b). On the concept of types: The type concept revisited. *American Anthropologist* 56: 42–53.

Gasche, H., and O. Tunca. (1983). Guide to archaeostratigraphic classification and terminology: Definitions and principles. *Journal of Field Archaeology* 10: 325–335.

Gould, S. J., and N. Eldredge. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* 3: 115–151.

Gould, S. J., and N. Eldredge. (1986). Punctuated equilibrium at the third stage. *Systematic Zoology* 35: 143–148.

Gould, S. J., and N. Eldredge. (1993). Punctuated equilibrium comes of age. *Nature* 366: 223–227.

Hooton, E. A. (1926). Methods of racial analysis. *Science* 63: 75–81.

Hughes, S. S. (1998). Getting to the point: Evolutionary change in prehistoric weaponry. *Journal of Archaeological Method and Theory* 5: 345–408.

Hull, D. L. (1970). Contemporary systematic philosophies. *Annual Review of Ecology and Systematics* 1: 19–54.

Hull, D. L. (1997). The ideal species concept—and why we can't get it. In M. F. Claridge, H. A. Dawah, and M. R. Wilson (eds.), *Species: The Units of Biodiversity*. London: Chapman and Hall, pp. 357–380.

Huxley, J. S. (1940). Toward the new systematics. In J. S. Huxley (ed.), *The New Systematics*. Oxford: Clarendon Press, pp. 1–46.

Kornbacher, K. D., and M. E. Madsen. (1999). Explaining the evolution of cultural elaboration. *Journal of Anthropological Archaeology* 18: 241–242.

Krieger, A. D. (1944). The typological concept. *American Antiquity* 9: 271–288.

LeTourneau, P. D. (1998). The "Folsom Problem." In A. F. Ramenofsky and A. Steffen (eds.), *Unit Issues in Archaeology: Measuring Time, Space, and Material*. Salt Lake City: University of Utah Press, pp. 52–73.

Lipo, C., M. Madsen, R. C. Dunnell, and T. Hunt. (1997). Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology* 16: 301–334.

Lyman, R. L., and M. J. O'Brien. (1997). The concept of evolution in early twentieth-century Americanist archeology. In C. M. Barton and G. A. Clark (eds.), *Rediscovering Darwin: Evolutionary Theory in Archeological Explanation*. Archeological Papers No. 7. Arlington, VA: American Anthropological Association, pp. 21–48.

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., and M. J. O'Brien. (1999). Americanist stratigraphic excavation and the measurement of culture change. *Journal of Archaeological Method and Theory* 6: 55–108.

Lyman, R. L., and M. J. O'Brien. (2000). Chronometers and units in early archaeology and paleontology. *American Antiquity* 65: 691–707.

Lyman, R. L., M. J. O'Brien, and R. C. Dunnell. (1997). *The Rise and Fall of Culture History*. New York: Plenum.

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.

Mayden, R. L. (1997). A hierarchy of species concepts: The denouement in the saga of the species problem. In M. F. Claridge, H. A. Dawah, and M. R. Wilson (eds.), *Species: The Units of Biodiversity*. London: Chapman and Hall, pp. 381–424.

Mayr, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.

Mayr, E. (1959). Darwin and the evolutionary theory in biology. In B. J. Meggers (ed.), *Evolution and Anthropology: A Centennial Appraisal.* Washington, DC: Anthropological Society of Washington, pp. 1–10.

Mayr, E. (1968). The role of systematics in biology. *Science* 159: 595–599.

O'Brien, M. J., T. D. Holland, R. J. Hoard, and G. L. Fox. (1994). Evolutionary implications of design and performance characteristics of prehistoric pottery. *Journal of Archaeological Method and Theory* 1: 259–304.

O'Brien, M. J., and R. L. Lyman. (1998). *James A. Ford and the Growth of Americanist Archaeology.* Columbia: University of Missouri Press.

O'Brien, M. J., and R. L. Lyman. (1999). *Seriation, Stratigraphy, and Index Fossils: The Backbone of Archaeological Dating.* New York: Kluwer Academic/Plenum.

O'Brien, M. J., and R. L. Lyman. (2000). *Applying Evolutionary Archaeology: A Systematic Approach.* New York: Kluwer Academic/Plenum.

O'Brien, M. J., R. L. Lyman, and R. D. Leonard. (1998). Basic incompatibilities between evolutionary and behavioral archaeology. *American Antiquity* 63: 485–498.

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.

Ramenofsky, A. F., and A. Steffen (eds.). (1998). *Unit Issues in Archaeology: Measuring Time, Space, and Material.* Salt Lake City: University of Utah Press.

Simpson, G. G. (1943). Criteria for genera, species, and subspecies in zoology and paleozoology. *New York Academy of Sciences, Annals* 44: 145–178.

Simpson, G. G. (1944). *Tempo and Mode in Evolution.* New York: Columbia University Press.

Simpson, G. G. (1945). The principles of classification and a classification of mammals. *American Museum of Natural History, Bulletin* 85.

Simpson, G. G. (1961). *Principles of Animal Taxonomy.* New York: Columbia University Press.

Spaulding, A. C. (1953). Statistical techniques for the discovery of artifact types. *American Antiquity* 18: 305–313.

Spaulding, A. C. (1954). Reply to Ford. *American Antiquity* 19: 391–393.

Spencer, C. S. (1997). Evolutionary approaches in archaeology. *Journal of Archaeological Research* 5: 209–264.

Sylvester-Bradley, P. C. (ed.). (1956). *The Species Concept in Palaeontology.* London: The Systematics Association.

Valentine, J. W., and C. L. May. (1996). Hierarchies in biology and paleobiology. *Paleobiology* 22: 23–33.

Whittaker, J. C., D. Caulkins, and K. A. Kamp. (1998). Evaluating consistency in typology and classification. *Journal of Archaeological Method and Theory* 5: 129–164.

Willey, G. R. (1953). Archaeological theories and interpretation: New World. In A. L. Kroeber, *Anthropology Today.* Chicago: University of Chicago Press, pp. 361–385.