A Wickliffe Perspective on Phase Definition
Kit W. Wesler
Department of Geosciences
Murray State University

Abstract

Ever since Willey and Phillips introduced phases, Mississippi Valley archaeologists have struggled to apply the concept consistently. Confusion resulted from the emphasis placed on the dimensions of assemblage, time, and space, effective restriction of many phase definitions to ceramic criteria, and difficulty in characterizing significant similarities among assemblages. Applying the Brainerd-Robinson coefficient of similarity to create an index of assemblage homogeneity, for which levels of significance can be calculated, offers a new perspective. Using assemblages from the Wickliffe Mounds, I suggest that we should expect sites within a phase to be no more homogeneous than samples from within a single site.

Today I’d like to dip my trowel into the churned soils of phase definition in the central Mississippi Valley. I do this with some trepidation, knowing that the subject has stirred some controversy, even acrimony. I should say that this is a preliminary formulation, so I hope you’ll bear with me.

The term “phase”, of course, was defined by Willey and Phillips in 1958 as “an archaeological unit possessing traits sufficiently characteristic to distinguish it from all other units similarly conceived... spatially limited to the order of magnitude of a locality or region and chronologically limited to a relatively brief interval of time.” Built into this definition was a classic formulation of the three dimensions of archaeological units: space, time, and form or content.

Theoretically, archaeological units can be based primarily on any of the three dimensions. Periods emphasize the chronological dimension, and hang spatial and cultural elements on the calendar. Willey and Phillips did not propose a taxonomy of temporal units. They did propose a taxonomy of spatial units, from site, the smallest, to locality, to region, to area. They suggested that a locality would be of a scale “occupied by a single community or a local group,” while a region would be “likely to coincide with minor physiographic subdivisions.”

In practice, I don’t think I’ve seen much use of the locality, but our practice of defining research areas by
physiographic features like drainages would seem to qualify as regions. Our use of such regions, however, seems not to reflect anything but physiography, and research within such a region usually investigates the entire archaeological record within it. In other words, we rarely use it as an archaeological unit.

Willey and Phillips’s most comprehensive discussion focused on the formal or content units: first component, then phase … but then jumps to “maximum units.” There is no fully-realized taxonomy: the phase is the working unit that matters to them.

So, there are three dimensions on which archaeological units can be defined, and although all units are informed by all three dimensions, there are three types of units that could be developed into hierarchical taxonomies. Temporal units—periods—would be defined first on time, then on content and space. Geographic units—localities, regions—would be defined first on space, then on content and time. Content units—phases—would be defined first on archaeological assemblages, then on time and spatial criteria.

Willey and Phillips plainly expected that content units would be the most practical and valuable units of analysis, and that the phase would be the level at which the analysis would be most productive. In fact, they made explicit the idea that material characteristics were of paramount importance: for a
phase, “the emphasis cannot be placed entirely on time;” a period “is determined by cultural criteria in the first place.”

To belabor the point: one culture sequence proposed for western Kentucky, Barry Lewis’s scheme, overtly placed the emphasis on the time dimension. He named several units based on a 200-year progression, and chose names to designate those units. However well those units may work for the analysis he wanted to do, by the Willey and Phillips definition they are periods, and many of us find it awkward to use the term “phase” in this case.

On the other hand, we might want to create a unit named, say, the Malden Plain phase, which would in effect be the Mississippian occupation of the Malden Plain. Unless we had formal criteria—traits of material culture—that distinguished the Mississippian of the Malden Plain from the Mississippian of the adjacent areas, to call this unit a phase would be a misnomer. It would be a region.

Most of us have accepted the basic idea of the phase as a formal unit, based on archaeological assemblages and traits, whose temporal and spatial dimensions would be defined by the limits of occurrence of these complexes of traits. The idea is fine. The practice has been somewhat less than consistent.

Michael O’Brien and Greg Fox have tackled this conundrum in several papers recently. In my view, what they are saying is
that phases in Southeast Missouri, in particular, have been defined somewhat loosely, without regular criteria to distinguish one phase from another. They are working around the problem of defining similarity among archaeological assemblages: how similar do assemblages have to be to belong to a phase? How different is an outlier? How do you measure similarity?

There are of course a number of papers that approach the problem of “Quantifying Diversity in Archaeology,” to quote the title of a 1989 book edited by Leonard and Jones. There are various techniques, borrowed from ecology, that measure something the archaeologists who use them call heterogeneity. Measures of heterogeneity combine measures of richness, having to do with the number of categories in the assemblage, and evenness, which has to do with to what extent the categories are equally represented within the assemblage. Evenness is a horror in archaeological pattern studies, where a few artifact categories—like Mississippian plain wares, or Stanley South’s Kitchen group in historical archaeology—tend to dominate assemblages and the real action is in small variations in low-count, low-percentage categories.

To these “concepts related to diversity” George Cowgill also added the notions of range, standardization, and uniformity of standardization. All of this can get pretty involved statistically, but I tend to agree with Dunnell, who called
these a set of measures in search of an application. For one thing, they haven't been applied to anything that really makes them look useful. More important for the phase problem, they measure heterogeneity within assemblages, not among assemblages. What we want to measure is similarity between assemblages, and homogeneity or heterogeneity within groups of assemblages.

O’Brien and Fox turned to the Brainerd-Robinson coefficient of similarity, which was derived specifically as a measure of similarity between assemblages where artifact counts are expressed as percentages. It measures on a scale of 200, with 200 signifying identical assemblages and 0 signifying completely different assemblages.

The Brainerd-Robinson coefficient was defined some time ago, but hit a wall in terms of application. It was proposed as a seriation device, but there are easier ways to do seriation, and we tend to think of seriation as an application to chronology. Seriation may reflect other dimensions than time, and a coefficient of similarity doesn’t specify what dimension of time, space or assemblage content is being measured, but the coefficient was typecast as a seriation device.

So Brainerd-Robinson coefficients are a technique in search of an application also, something that a few people like George Cowgill and Albert Spaulding seem to dust off now and then for heuristic purposes but haven’t applied very usefully.
Parenthetically, historical archaeologists apparently haven’t even thought about Brainerd-Robinson because it was derived for prehistoric sites, and historical archaeologists like to think of their field as something segregated from prehistory—but that’s a diatribe for another time.

Also parenthetically, George Cowgill explained cogently why the Brainerd-Robinson coefficient is more useful for archaeology than other statistics, like Pearson’s r, in a 1990 article in *American Antiquity*. In exploring the similarity measurement problem, I also found that most of the standard statistical approaches to social science data didn’t fit our needs very well. To use the raw data—assemblage counts—requires fairly consistent sample sizes, which we rarely have in archaeological assemblages. To allow for sample size biases, we either use percentages or do some fairly involved, and I think dubious, statistical manipulation. But the statistical packages don’t apply very well to percentage data, for the most part. The Brainerd-Robinson coefficient works with percentage data.

In considering phase definition, O’Brien and Fox applied Brainerd-Robinson to a new purpose: showing the wide variability of collections from the same site, to make a point about sample biases due to different collection techniques. Here I can point out that historical archaeologists noted patterned differences
between surface collected and excavated assemblages a couple of decades ago.

slide  O’Brien and Fox also used Brainerd-Robinson to measure relationships between assemblages in order to define groups of more- and less-like assemblages, that is, potentially, phases. I like the approach, but I have two problems: first that, like just about everybody, they use only ceramic type frequencies, and second, that they are fairly arbitrary about what the threshold coefficient of similarity is. I don’t see a discussion of what is a significant level of similarity.

I think there’s a further step we can take towards defining significant similarity. If we have comparable data collection techniques, we can use Brainerd-Robinson to measure assemblage similarity. We can then use very simple statistics to look at the distributions of similarity coefficients among groups of assemblages: this becomes an index of homogeneity, and a way to assess whether a single or a group of assemblages is different from others.

For instance, Greg Fox calculated a Brainerd-Robinson coefficient matrix for Stephen Williams’s analytical units from the Crosno site. Taking the coefficients as data points, we can calculate quickly that they average 185.06, with a standard deviation of 7.38. I’ll call this the Brainerd-Robinson Index of Homogeneity: for Crosno, $185.06 \pm 7.38$. The coefficients are
quite high, and the distribution is tight, indicating a great deal of mutual similarity. In fact the test units are in close proximity to each other and I suggest that the occupation time represented is relatively short, so we can expect a high degree of internal homogeneity among the analytical units. But—by what standard can we accept this as a high degree of similarity? And can we generalize this or any index level as a similarity justifying inclusion into a phase?

**slide** Here is where I bring in the Wickliffe data. The Wickliffe site is a Middle Mississippian village on the bluffs of the Mississippi River at the western edge of Kentucky, occupied by Native Americans from about AD 1100 to 1350. Through dedicated determination or lack of imagination, I put fourteen summers of test excavations into the site. The test units represent a spatial sample, not statistically random but without intentional bias, throughout the major sectors of the site. Based on ceramic and stratigraphic criteria, I define an Early, Middle, and Late Wickliffe sequence within the 250 year occupation, and I can treat the three internal periods as closely-related components. For the following data, I have used midden deposits, not mound fill or feature deposits.

**slide** I constructed Brainerd-Robinson matrices for the ceramic assemblages for each period, including only those samples with more than 25 sherds. I used 12 types or type groups for this
analysis. I set up the calculations, by the way, in Microsoft Excel, with a template for 65 assemblages and 12 variables. I stopped there, frankly, because I got bored out of my skull filling in all those cells to do the calculations. Anyone who’d like to play with these statistics is welcome to email me and I’ll attach the Excel template file.

For Early Wickliffe middens, 14 samples, the index of homogeneity is 178.67 ± 9.39. For 27 Middle Wickliffe samples, the index is 182.86 ± 7.8, and for 52 Late Wickliffe samples, the index is 187.63 ± 6.19. There is a slight tendency for the deposits to become more homogeneous through time. More interesting, these statistics are quite close to the index computed earlier for Crosno, 185.06 ± 7.38, which I consider to be analogous to a Late Wickliffe assemblage.

As O’Brien and Fox noted, one factor in these high similarity coefficients is the overwhelming effect of plain wares. Mississippi Plain tends to be 80% of the assemblages or more, and Bell Plain often another 10%. O’Brien and Fox recalculated many of their samples after deleting the plain wares. I found that if I deleted both Mississippi Plain and Bell Plain, my sample sizes plummeted, so I retained the Bell Plain. I also kept the grog tempered category, since there is a small presence of grog-tempered ceramics throughout the
Mississippian sequence in western Kentucky. I distinguished between two varieties each of Matthews Incised and O’Byam Incised, but dumped the “other” category to get back to 12 variables.

*slide* With these assemblages, the indices of homogeneity recalculated to: Early Wickliffe $158.05 \pm 21.37$; Middle Wickliffe $160.48 \pm 13.91$; and Late Wickliffe $164.84 \pm 19.44$. The temporal trend is the same. The standard deviations are larger, but the index values are still in the upper quarter of the Brainerd-Robinson range, and the ranges overlap the first set at the one-sigma interval.

*slide* Since I know you’re asking, I also calculated Brainerd-Robinson matrices for the Early, Middle and a sample of Late Wickliffe assemblages together. The results were quite comparable to the indices for Early Wickliffe. The variability among the three periods is no more than the variability within the Early and Middle periods.

I can go in a couple of directions here. For one thing, I can look, within each group of samples, at those deposits that have similarity coefficients below one or two standard deviations of the average, and pick out anomalous test units to see what makes them special.
I can also start to construct similarity matrices for regional sites, which gets back to the problem of phase definition. There is a potential problem here: comparability of data collection techniques. Field techniques are fairly standardized, and screen size is usually specified in reports. But lab procedures may not be quite as standard. I sort ceramics on a half-inch screen, and ignore the smaller sherds. Many reports specify the number of sherds smaller than a threshold, which can be accounted for, but some do not. Also, many reports specify major type categories, but when describing the “other” sherds, do not attribute all of them to units or excavation levels, which can make statistics difficult.

As an experiment, I compared my Wickliffe totals to site totals for other reported sites in western Kentucky. The University of Illinois’s Western Kentucky project, directed by Barry Lewis, reported tests excavations at eight Late Woodland and Mississippian sites, and David Pollack and Jimmy Railey reported the Mississippian Chambers site. Only the Turk site reported sherds level-by-level, so that I could organize the data into assemblage groups comparable to Early, Middle and Late Wickliffe. The other sites’ data were reported by test unit, conflating internal sequences.

When I calculated the Brainerd-Robinson index for all of the excavated ceramics from all of the sites and components, the
result was 109.76 ± 58.86. This is a low level of average similarity, and a very wide standard deviation, compared to the intra-Wickliffe data set. This is not surprising: surely we would all agree the Late Woodland and Mississippian sites in a region belong to different phases.

Isolating the substantially Mississippian sites produces an index value of 160.08 ± 24.26, much closer to the intra-Wickliffe levels of similarity and overlapping in the one-sigma ranges. In other words, the variation among these sites is not much greater than the variation within the Wickliffe site. Two sites, Adams and Twin Mounds, have substantial Late Woodland components, and I suspect that if we could separate the Mississippian from the Late Woodland deposits in these two sites, the Mississippian component index of homogeneity would be quite comparable to the intra-Wickliffe index value. Would this measure, then, give us an operating significance level for an index of homogeneity for defining a phase?

I only have time to note the other problem with phase definition, that the Willey and Phillips model never, I don’t think, envisioned a reliance solely on ceramic typology for defining phases. We need to look at other assemblage characteristics, as well. I have played with an idea of slide assemblage patterning borrowed from Stanley South, and organized the Wickliffe assemblages into Ceramics, Arms, Personal,
Activities, and Debitage groups, which show some patterning within the Wickliffe site. Intra-site indices of homogeneity for these artifact groups are at comparable levels to the ceramics indices. I can also look at vessel form assemblages. There must be other assemblage variables worth looking at—faunal, chert sources, etc. The problems of lab and reporting comparability among projects become critical here, but for studies of intra-project patterning, the method may yield some interesting results.

To conclude, I’d like to make several suggestions.

First, we need to explore assemblage variation in a number of ways, and not designate one type component and try to create phases based on unrigorous comparisons to that assemblage.

Second, we need to build variability in different kinds of assemblages, not just ceramic type frequencies, into phase definitions.

Third, I suggest that sites within a phase should be no more homogeneous than samples within a complex site. The Wickliffe data may offer an initial standard for variability within a site, as the site with the most samples.

Whether the Brainerd-Robinson index or some other technique will be the most useful approach to this problem, I don’t know yet, but at least it’s a fun way to play with a lot of numbers.

Thank you.