

Sampled to Death? The Rise and Fall of Probability Sampling in Archaeology

Edward B. Banning 

After a heyday in the 1970s and 1980s, probability sampling became much less visible in archaeological literature as it came under assault from the post-processual critique and the widespread adoption of “full-coverage survey.” After 1990, published discussion of probability sampling rarely strayed from sample-size issues in analyses of artifacts along with plant and animal remains, and most textbooks and archaeological training limited sampling to regional survey and did little to equip new generations of archaeologists with this critical aspect of research design. A review of the last 20 years of archaeological literature indicates a need for deeper and broader archaeological training in sampling; more precise usage of terms such as “sample”; use of randomization as a control in experimental design; and more attention to cluster sampling, stratified sampling, and non-spatial sampling in both training and research.

Keywords: probability sampling, statistics, survey, history of archaeology, archaeological pedagogy

Después de un apogeo en los años setenta y ochenta, el muestreo probabilístico se hizo mucho menos visible en la literatura arqueológica, ya que se vio amenazado por la crítica posprocesal y la adopción común de la “encuesta de cobertura completa”. Después de 1990, la discusión publicada sobre el muestreo probabilístico rara vez se desvió de los problemas del tamaño de la muestra en los análisis de artefactos, restos de plantas y animales, mientras que la mayoría de los libros de texto y el entrenamiento arqueológico limitaron el muestreo al estudio regional e hicieron poco para equipar a las nuevas generaciones de arqueólogos con este aspecto crítico de diseño de la investigación. Un resumen de los últimos 20 años de literatura arqueológica indica la necesidad de una formación arqueológica más profunda y amplia en el muestreo, el uso más preciso de términos como “muestra”, el uso de la aleatorización como control en el diseño experimental y una mayor atención al muestreo conglomerado, muestreo estratificado, y muestreo no espacial tanto en capacitación como en investigación.

Palabras clave: muestreo de probabilidad, estadísticas, prospección, historia de la arqueología, pedagogía arqueológica

Recently, I was asked to write a contribution on spatial sampling in archaeology (Banning 2020), with case studies to illustrate best practices. To my surprise, I had difficulty finding any examples, let alone best practices, of probability sampling—spatial or otherwise—in archaeological literature of the last 20 years, aside from Orton’s (2000) excellent book. Given that sampling theory is a critical aspect of research design and control for bias, this puzzled and concerned me.

There can be legitimate reasons not to employ probability sampling. What rattled me when I tried to find those case studies is the possibility that many archaeologists are neglecting probability sampling for the wrong reasons.

Here, I explore some possible reasons for this neglect before offering some suggestions for restoring formal sampling to a substantive role in our practice. First, however, let us review briefly the purpose and nature of probability sampling, and a brief history of its use in archaeology.

Edward B. Banning (ted.banning@utoronto.ca, corresponding author) ■ Department of Anthropology, University of Toronto, 19 Russell St., Toronto, Ontario, M5S 2S2, Canada

American Antiquity 86(1), 2021, pp. 43–60

Copyright © 2020 by the Society for American Archaeology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

doi:10.1017/aaq.2020.39

What Is Sampling?

Sampling entails two important concepts: (1) the population, or set of phenomena—such as sites, features, spaces, artifacts, bone, and charcoal fragments—whose characteristics are of interest, and (2) the sample, a subset of the population that we actually examine. Our interest in the sample is that it might tell us something about the population. In archaeology, we often have populations that consist of spaces, such as excavation squares, because we cannot enumerate populations of sites, artifacts, or “ecofacts” that we have not yet surveyed or excavated. A “sampling frame” is a list of the population’s “elements” or members, or a grid or map for identifying the set of spatial elements in a spatial population. Sample size is just the number of elements in the sample, whereas sampling fraction is the sample size divided by the number of elements in the whole population, whether known or not. Even archaeologists who do not formally employ sampling theory accept that a large sample is a better basis for inferences than a very small sample.

Archaeologists sample all the time, if only because cost or other factors make examination of whole populations impractical or unethical. We also recognize that taphonomic factors can distance our sample of a “fossil assemblage” still farther from a “deposited assemblage” that may be our real population of interest (Holtzman 1979; Meadow 1980). The question is, How confident should we be about inferences based on a small subset of a population?

For some kinds of samples, not very. “Convenience” or “opportunistic” samples are just the sites, artifacts, or plant or animal remains that come to hand, often because they are already sitting in some lab. Potentially better are “purposive samples” that result from conscious selection of certain members of the population because of the perception that they provide superior information for some purpose, such as excavation areas selected for their probability of yielding a long, stratified sequence. Samples such as these are not flawed, for certain purposes at least, but they entail the risk that they may not be “representative” of the population of interest. In other words, the sample’s characteristics might not be very similar to the characteristics

of the whole population. A nonrandom difference between the value of some population characteristic (statisticians call this a “parameter”) and the value of that characteristic (or “statistic”) in a sample is “bias.”

Probability sampling is a set of methods with the goal of controlling this risk of bias by ensuring that the sample is “representative” of the population so that we can estimate a parameter on the basis of a statistic. This always involves some randomness. The classic probability sampling strategies include simple random sampling with replacement, in which every “element” of the population—whether an artifact, bone fragment, space, or volume—has an equal probability of selection at each and every draw from the population. This is like picking numbers from a hat but then replacing them so that some elements can be selected more than once. Alternatively, we may remove elements once they are randomly selected (random sampling without replacement) so that the probability of selection changes as sampling progresses and no element is selected more than once. Another is systematic sampling, in which we randomly select the first element and then all the others are strictly determined by a “spacing rule.” For example, we might organize artifacts in rows, randomly select one of the first four artifacts by rolling a die (ignoring 5 and 6), and then take every fourth artifact in sequence to yield a 25% sampling fraction. Stratified sampling involves dividing the population into subpopulations (“strata”) that differ in relevant characteristics before sampling within them randomly or systematically. Systematic unaligned sampling is a specifically spatial design meant to ensure reasonably even coverage of a site or region without as much rigidity as a systematic sample (Figure 1). Most probability sampling designs are variations or combinations of these basic ones.

Sample elements need not be spatial (Figure 2), but the fact that archaeologists can rarely specify populations of artifacts or “ecofacts” in advance often forces them to employ cluster sampling. Cluster samples occur whenever the population of interest consists of items such as artifacts, charcoal, or bone fragments, but the population actually sampled is a spatial one, such as a population of 2×2 m squares

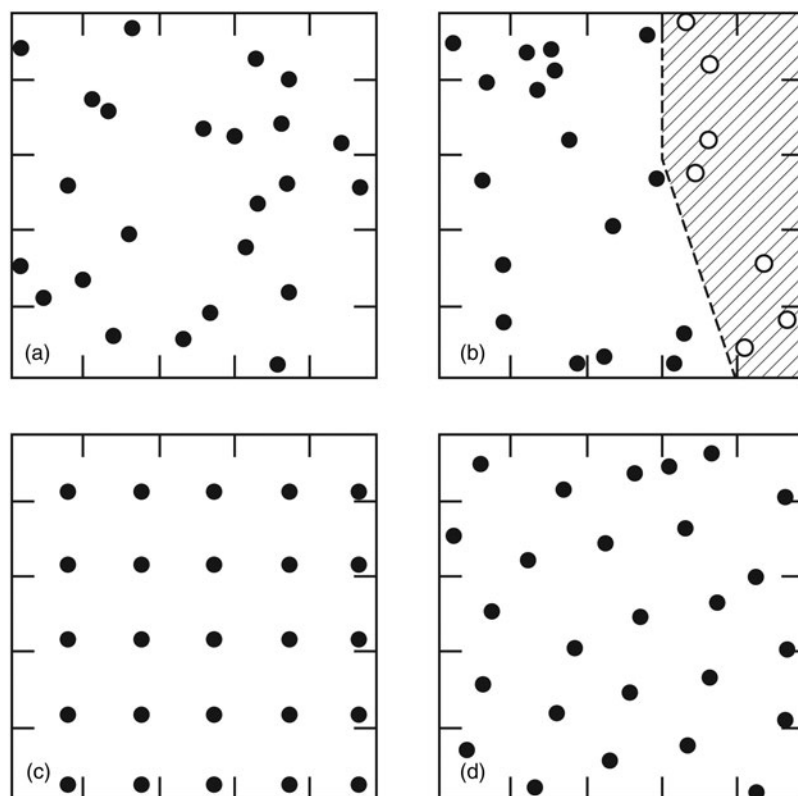


Figure 1. Hypothetical examples of some spatial sampling designs (after Haggett 1965:Figure 7.4) that were repeated in dozens of later archaeological publications: (a) simple random, (b) stratified random, (c) systematic, and (d) systematic unaligned.

(Mueller 1975a). Cluster samples require statistical treatment that differs from that for simple random samples (Drennan 2010:244–246; Orton 2000:212–213) because of the phenomenon called “autocorrelation,” which is that observations that are close together are likely to be more similar to one another than ones that are far apart. In the case of lithics, it is likely that multiple flakes found near each other came from the same core, for example.

Multistage sampling is a variety of cluster sampling in which there is a hierarchy of clusters. For example, we might first make a stratified random selection of sites that have been excavated, then randomly select contexts or features from the selected excavations (themselves generally already samples of some kind), then analyze the entire contents of the sampled contexts.

Another important type is Probability Proportional to Size, or PPS sampling (Orton 2000:34).

This involves randomly or systematically placed dots or lines over a region, site, thin section, pollen slide, or other area. Only sites, artifacts, or mineral or pollen grains that the dots or lines intersect are included in the sample. Because the dots or lines are more likely to intersect large items than small ones, it is necessary to correct for this effect to avoid bias.

In general, probability sampling is preferable to convenience or purposive sampling whenever we should be concerned whether or not the sample is representative of a population—and, consequently, suitable for making valid inferences about it. Convenience sampling is acceptable for some clearly defined purposes when probability sampling is impossible or impractical, and purposive sampling can be preferable when we have very specific hypotheses whose efficient evaluation requires targeted, rather than randomized, observations.

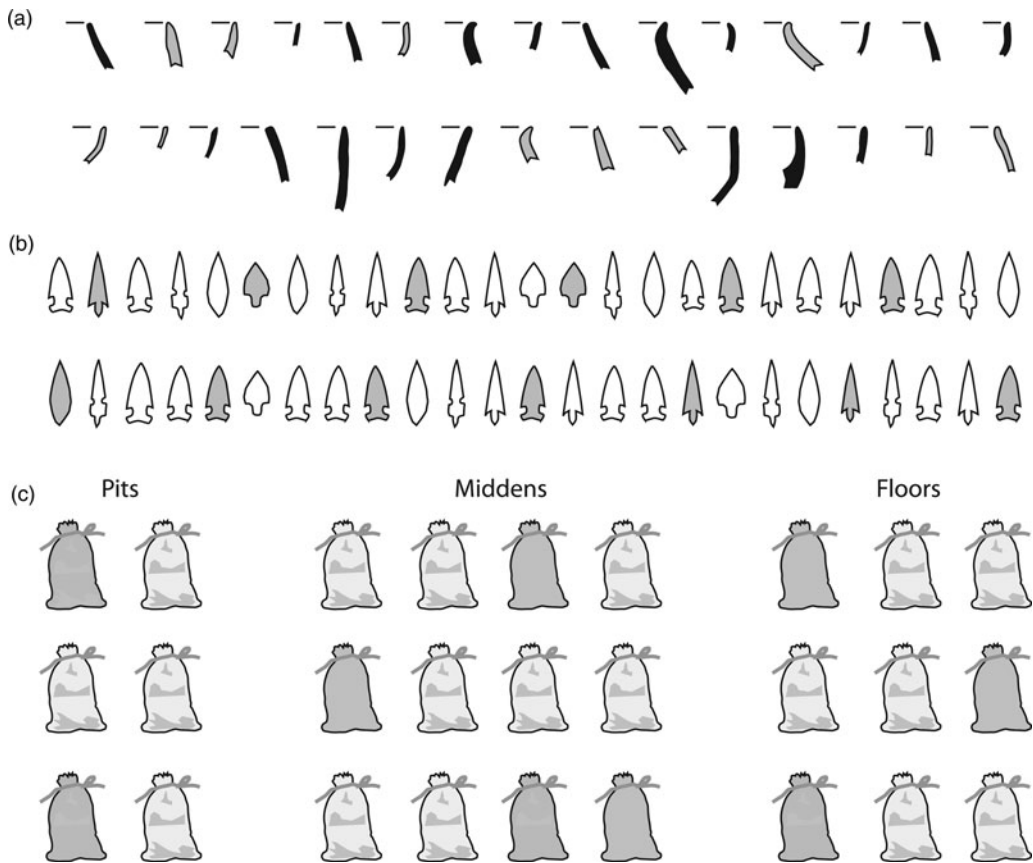


Figure 2. Some hypothetical examples of nonspatial samples, with selected elements in gray: (a) simple random sample of pottery sherds (the twelfth sherd selected twice), (b) 25% systematic sample of projectile points arranged in arbitrary order, and (c) stratified random sample of sediment volumes for flotation.

The Rise of Archaeological Probability Sampling

What was it that once made sampling theory appeal to archaeologists? Its perception as “scientific,” no doubt, was a contributing factor. As the previous section suggests, a better incentive was that, by controlling sources of bias, it permits valid conclusions about populations when observing entire populations is impossible, undesirable, wasteful, or unethical. Sampling allows us to evaluate the strength of claims about populations with less worry that results are due to chance or, worse, our own preconceptions (Drennan 2010:80–82; Orton 2000:6–9).

Some of the earliest attention to sampling in archaeology concerned sample size. Phillips and colleagues (1951) made frequent reference

to samples of sites and pottery, and especially the adequacy of sample sizes for seriation. In one instance, they even drew a random sample of sherds (1951:77). They did not, however, employ formal methods to decide what constituted an “adequate” sample size, and sampling did not attract much explicit attention from archaeologists until the 1960s (Rootenbergh 1964; Vescelius 1960).

Binford (1964) was particularly influential in archaeologists’ adoption of probability sampling, presenting it as a key element of research design. He summarized the main sampling strategies reviewed in the last section and identified different kinds of populations and the role of depositional history in their definition. He also recognized that confounding factors—such as vegetation, construction, land use, and

accessibility—could complicate sampling designs and inferences from spatial samples.

He also inadvertently fostered some misconceptions. Despite advocating nuanced decisions on sample size earlier in the article, Binford dismissed attention to sample size as “quite complicated,” and continued with, “for purposes of argument, . . . we will assume that a 20% areal coverage within each sampling stratum has been judged sufficient” (1964:434). Although this was just a simplifying example, later archaeologists often took 20% as a recommended sampling fraction. Similarly, some archaeologists seem to have taken his mention of soil type as grounds for stratification as received wisdom, and they used soil maps to stratify spatial samples whether or not this made sense. Despite his assertion that probability sampling should occur “on all levels of data collection” (1964:440), both this article and much of the literature it inspired strongly privilege sampling in regional surveys, with less attention to sampling sites, assemblages, or artifacts (but see Orton 2000).

Soon, sampling appeared in texts used to educate the next generation of archaeologists (Ragir 1967; Watson et al. 1971). Most focused on the basic spatial sampling designs. Generally lacking was discussion of when probability sampling was appropriate and how to define populations or plan effective stratified or cluster samples.

An outpouring of literature on sampling in regional survey (e.g., Cowgill 1970, 1975; Judge et al. 1975; Lovis 1976; Williams et al. 1973), surface collection (Redman and Watson 1970), excavation (Hill 1970), zooarchaeology (Ambrose 1967), and artifact analysis (Cowgill 1964) also appeared. There were more general reviews (Mueller 1975b; O’Brien and Lewarch 1979; Redman 1974) and desktop simulations of sampling designs (Mueller 1974; Plog 1976). Sampling soon saw application outside North America (e.g., Cherry et al. 1978; MacDonald et al. 1979; Redman and Watson 1970), and the number of articles in *American Antiquity* that discussed or used probability sampling grew rapidly until 1980 (Figure 3).

Then, the literature shifted to more focused topics, such as determining sample sizes (Dunnell 1984; Leonard 1987; McManamon 1981; Nance 1981), ensuring that absences of

certain classes are not due only to sampling error (Nance 1981), or sampling shell middens (Campbell 1981).

During the 1970s, much research still used purposive sampling or ignored this sampling wave. Even authors who did not embrace sampling, however, tended to be somewhat apologetic, offering caveats about their samples’ usefulness or describing their research as preliminary.

The Fall of Probability Sampling in Archaeology

Articles published in *American Antiquity* since 1960 and *Journal of Field Archaeology*, two long-lived journals that regularly publish articles on archaeological methods (Figure 3 and Supplemental Text 1), show that substantive discussion or mentions of sampling in the statistical sense peaked about 1980 in the former and 1990 in the latter, then declined, albeit with some recovery in the late 1990s and again in the last few years, never returning to pre-1990 levels. What these graphs do not reveal is that articles prior to 1985 tend to be *about* sampling, while most after 1990 just *mention having used* some kind of random or systematic sample, usually without presenting any details. Those few about sampling in the later period usually pertain to sample-size issues in zooarchaeology and paleoethnobotany rather than to research design more generally. It is hard to imagine that there was nothing further to say about probability sampling or that it had become too routine to warrant comment, especially as research in statistics developed considerably after 1970 (e.g., Orton 2000:11; Thompson and Seber 1996).

Remarkably, a common claim of the 1990s is that some pattern “is highly unlikely to result from sampling error or random chance” (e.g., Falconer 1995:405), despite relying on small or non-probabilistic samples. Other authors acknowledge bias in their data but go on to analyze them as though they are unbiased, or describe “sampling designs” expected to provide representative samples by standardizing sampling elements without reference to probability sampling (e.g., Bayman 1996:407; Walsh 1998:582).

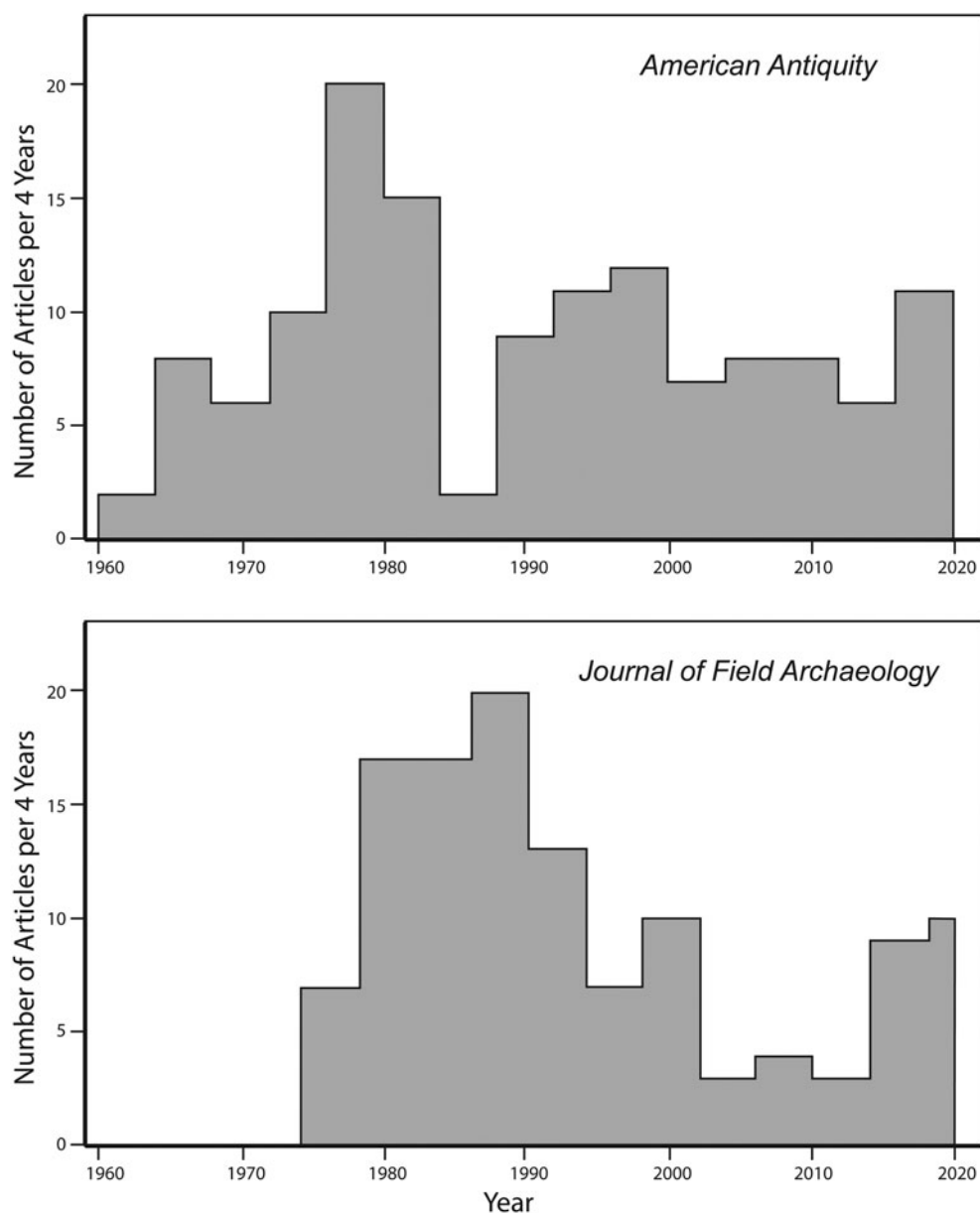


Figure 3. The frequency of articles with substantive discussion of sampling or based at least partly on explicit probability samples in *American Antiquity* (1960–2019) and *Journal of Field Archaeology* (1974–2019). Note that there was an interruption in *Journal of Field Archaeology* from 2002 until early 2004 and that 2018–2019 have five articles (10 per four years).

A blistering critique (Hole 1980) that exposed flaws in then-recent archaeological sampling—including arbitrary sampling fractions, the suppression of standard error through impractically small sample elements, and the ignoring of prior information—foreshadowed this decline.

Hole (1980:232), however, was not criticizing sampling per se, just its misapplications. Even though some authors judged her critique as extreme (Nance and Ball 1986; Scheps 1982), the fact that none of the 17 articles that cited it from 1981 to 1999, according to Google Scholar,

advocated abandoning probability sampling suggests that it had little or no role in decreasing interest or expertise in sampling. The following sections consider more likely candidates.

The Post-Processual Critique

During the 1980s, attacks on sometimes pseudo-scientific or dehumanizing examples of New Archaeology engendered an anti-science rhetoric that may have made probability sampling a victim. Shanks and Tilley led this attack by arguing that mathematical approaches entail assumptions that theory is value free, and that “categories of analysis are necessarily designed to enable certain calculations to be made” (1987:57). McNany and Rowe (2015) explicitly connect rejection of probability sampling with the post-processual paradigm. More recently, Sørensen (2017) argues against a new “scientific turn” that devalues the humanities and fetishizes “scientific facts.” What he criticizes explicitly, however, is not really the use of samples but the use of inadequate ones (Sørensen 2017:106). This is not a problem with science; it just underscores the need for better sampling.

Furthermore, adherents of the interpretive paradigm still base inferences on samples and analyze data as though they represent something more than the sample itself. Even Shanks and Tilley (1987:173–174) explicitly used stratified sampling in their analysis of beer cans and based bar graphs and principal components analysis on this sample (Shanks and Tilley 1987:173–189; see also Cowgill 1993; VanPool and VanPool 1999; Watson 1990).

Similarly, Shanks (1999) relies on the quantitative distribution of motifs in a “sample of 2,000 Korinthian pots” (1999:40). This is an opportunistic sample of “all complete pots known” to Shanks, but he expects them to represent populations of artifacts and the people who made them, such as the pottery of “archaic Korinth” (Shanks 1999:2, 9, 10, 151). He further generalizes about the “early city state,” an even more abstract population (Shanks 1999:210–213), and wonders if his sample is “somewhat biased” for some purposes (1999:41).

Apparently, formal sampling and generalization from sample to population are not

incompatible with interpretive archaeology. Just as “atheoretical” archaeologists inescapably use theory (Johnson 1999:xiv, 6–8), avowedly anti-science archaeologists still use statistical reasoning and sampling. Their anti-science rhetoric, however, may still have had a chilling influence on explicit archaeological sampling.

The “Full-Coverage” Program

When Fish and Kowalewski (1990) published *The Archaeology of Regions*, “full-coverage survey” had already begun to trend. Its premise that small samples are an inadequate basis for some kinds of research is undeniable (Banning 2002:155): a small sample can never capture all nuances of a settlement system and suffers from “the Teotihuacan effect”—the risk of omitting key sites in a settlement system (Flannery 1976:159). The solution, according to most authors in this volume, is to survey an entire landscape at somewhat consistent intensity.

Its classic example is the Valley of Mexico survey. Rather than only examining a subset, surveyors examined every “accessible” space within a “universe” of spatial units with pedestrian transect intervals ranging from 15 to more than 75 m, but typically 40–50 m (Parsons 1990:11; Sanders et al. 1979:24).

Kowalewski (1990) identifies the main advantages of full coverage, claiming that it captures greater variability and larger datasets, facilitates analysis of spatial structure, and is better at representing rare observations. He also highlights its flexibility of scale in that it does not force researchers to “lock in” to an analytical unit size (cf. Ebert 1992), and he correctly notes that much archaeological research is *not* about parameter estimation.

The discussants who close out *The Archaeology of Regions*, however, were not as convinced that full coverage was better than sampling. With surveyor intervals as large as 100 m, most or all “full-coverage surveys” were actually still sampling, potentially having missed even easily detectable sites with horizontal dimensions less than the transect interval. These are really systematic transect samples, and PPS samples of sites, whose main virtue is even and somewhat consistent coverage (Cowgill

1990:254; Kintigh 1990:238). “Full coverage” does not mean anything close to 100% coverage unless transect intervals are extremely small and visibility and preservation are excellent (Given et al. 1999:22; Sundstrom 1993).

Fred Plog (1990) specifically rebuts many of Kowalewski’s claims, noting that well-designed stratified samples capture variability well, whereas volume of data is correlated with survey effort, irrespective of method. The claim that full-coverage surveys perform better at capturing rare sites is true only for large, obtrusive ones, not ones that are small or unobtrusive. As most full-coverage surveys really use transects as spatial units, they are also “locked in” to their transect spacings.

Due to the fact that most full-coverage surveys are systematic PPS samples, they yield biased estimates of some parameters—such as mean site size, the proportion of small sites, and the rank-size statistic—because they underrepresent small, unobtrusive sites and artifact densities unless their practitioners correct for this (Cowgill 1990:252–258). None of the surveys in *The Archaeology of Regions* did so, however.

In the aftermath of this book and a session decrying sampling at the 1993 Theoretical Archaeology Group conference, “almost everybody was against sampling” (Kamermans 1995:123). The “brief flirtation with survey sampling” led to “consensus . . . that best practice involves so-called ‘full-coverage’ survey” (Opitz et al. 2015:524). Despite its focus on spatial sampling, this probably influenced attitudes to sampling more generally.

Misunderstanding Sampling

Certain misconceptions also discouraged interest in sampling. Binford (1964) had proclaimed that sampling requires populations of equal-sized spatial units. Many archaeologists found that arbitrary grids, especially of squares, were rarely practical or useful because their boundaries did not correspond with meaningful variation on the ground. Sampling universes, however, do *not* have to consist of any particular kind of spatial unit (Banning 2002:86–88; Wobst 1983). Even as probability sampling was in its early decline, some projects successfully used sample

elements that conformed to geomorphological landforms, field boundaries, or urban architectural blocks (e.g., Banning 1996; Kuna 1998; Wallace-Hadrill 1990).

Some archaeologists worried that fixed samples fail to include rare items or represent diversity accurately. Others, however, found solutions. One is to supplement a probability sample with a purposive one (Leonard 1987; Peacock 1978); it is not appropriate to combine the two kinds of samples to calculate statistics, but researchers can use the probabilistic data to make parameter estimates for common things and the purposive sample to characterize rare phenomena or establish “detection limits” on their abundance. Another is to use sequential sampling instead of a fixed sample size (Dunnell 1984; Leonard 1987; Nance 1981; Ullah et al. 2015). This involves increasing sample size until some criterion is met, such as a leveling off in diversity or relative error.

Finally, some archaeologists have the mistaken idea that sampling is a way to find sites. Spatial sampling is actually rather poor at site discovery, but this does not discount its suitability for making inferences about populations (Shott 1985; Welch 2013). That sampling does not ensure site discovery is not a good reason to abandon it.

Opportunity and Exchangeability

The ubiquity of opportunistic populations and samples in archaeology may also discourage interest in formal sampling. In heritage management, for example, the “population” often corresponds to a project area that depends on development plans rather than archaeological criteria. In a corridor survey for a pipeline, a project area could intersect a large site, yielding a sample of cultural remains that may or may not be representative, with little opportunity even to determine the site’s size or boundaries, except in jurisdictions that offer some flexibility (e.g., Florida Division of Historical Resources 2016:14; and see below).

But this is not unique to cultural resource management (CRM). Archaeologists frequently treat an existing collection as a population, and they either sample it or study it in its entirety.

Biases could result from the nature of these opportunistic samples, but that does not mean we cannot evaluate their effects (Drennan 2010:92). Accompanying documentation might even facilitate an effective stratified sample.

Bayesian theory potentially offers some respite through its concept of exchangeability (Buck et al. 1996:72–78). An opportunistic sample may be adequate for certain purposes, as long as relevant observations on the sample are not biased with respect to those purposes, even if we can expect bias with regard to other kinds of observations. A collection of Puebloan pottery formed in the 1930s, or acquired by collectors, might include more large, decorated sherds or higher decorative diversity than would the population of sherds or pots in the site of origin because of the collectors' predispositions. Such a sample would provide biased estimates of the proportion of decorated pottery, but it might be acceptable for estimating the proportions of temper recipes in pottery fabrics, for example.

However, there is no reason to think that Bayesian exchangeability has any role in archaeologists' attitudes to probability sampling. Few texts on archaeological analysis even mention exchangeability (Banning 2000:88; Buck et al. 1996:72–74; Orton 2000:21). One other does, but without naming it (Drennan 2010:88–92). In archaeological research literature, I was only able to find a single example (Collins-Elliott 2017). Clearly, those who have eschewed probability sampling have not been aware of this concept.

Undergraduate Statistical Training

Another candidate cause for the decline is archaeological training (cf. Cowgill 2015; Thomas 1978:235, 242). A publication on teaching archaeology in the twenty-first century (Bender and Smith 2000) ignores sampling design, or even statistics more generally, except for a single mention of sampling as a useful skill (Schuldenrein and Altschul 2000:63). A proposed area for reform of archaeological curriculum, "Fundamental Archaeological Skills," is silent on research design, sampling, and statistics except to list statistics as a "basic skill" in graduate programs (Bender 2000:33, 42). At least one article

on archaeological pedagogy mentions, but does not develop, the role of sampling in field training (Berggren and Hodder 2003).

A review of undergraduate textbooks leads to some interesting observations. The selection (Supplemental Text 2) includes all of the English-language textbooks I could find that cover archaeological methods, but it limits multiple-edition books to the earliest and latest editions I could access.

Renfrew and Bahn's (2008:80–81) explication of major spatial sampling strategies is typical. After briefly mentioning non-probabilistic sampling, they describe random, stratified random, systematic, and systematic unaligned sampling designs, all in spatial application. They illustrate these with the same maps (from Haggett 1965:Figure 7.4) as has virtually every archaeology text that describes sampling since Stephen Plog (1976:137) used them in *The Early Mesoamerican Village* (Figure 1). For stratified sampling, they do not mention the rationale for strata or the need to verify that strata are statistically different. As usual, sampling's justification is that "archaeologists cannot usually afford the time and money necessary to investigate fully the whole of a large site or all sites in a given region" (Renfrew and Bahn 2008:80), while they say that probability sampling allows generalizations about a site or region.

Most texts that do not specialize in quantitative methods give, at best, perfunctory attention to sampling. Not one of 54 introductory texts in my list mentions sequential sampling (Dunnell 1984) or the newer development, adaptive sampling (Thompson and Seber 1996), and only five make any mention of sample size, sampling error, or nonspatial sampling. All but two only present sampling in regional survey, and only four mention alternatives to geometrical sample elements. A few misrepresent sampling as a means to *find* things, especially sites, rather than to estimate parameters or test hypotheses (e.g., Muckle 2014:99; Thomas 1999:127). Seventeen more specialized texts provide a fuller discussion of sampling, but they probably reach smaller, more advanced audiences.

Turning to curriculum, explicit statistical requirements are far from universal. Although

variations in how to describe undergraduate programs make comparison difficult, I was able to find information online for 24 of the 25 most highly ranked undergraduate programs internationally (Supplemental Text 3; Quacquarelli Symonds 2019). Of these, at least five (21%) require study in statistics, and eight (33%) have in-house statistical or data-analysis courses. Some may include statistics in other courses, such as science or laboratory courses. At least five have courses on research design (four have courses that may include some research design), but it is unclear whether these cover sampling. Many programs have an honors thesis or capstone course that could include sampling.

Although some archaeology programs emphasize quantitative methods, one of which even says that “quantitative skills and computing ability are indispensable” (Stanford University 2019), the overall impression is that knowledge of statistics or sampling is optional. The only article I found that explicitly addresses sampling education in archaeology (Richardson and Gajewski 2002) is not by archaeologists but by statisticians in a journal on statistics pedagogy.

Publication and Peer Review

One might also ask why the peer-review process does not lead to better explication of sampling designs. As one anonymous reviewer of this article pointed out, this is likely due to a lack of sampling and statistical expertise among a significant proportion of journal editors and manuscript reviewers who, after all, probably received training in programs much like the ones reviewed in the last section.

Archaeological Sampling in the Twenty-First Century

As the histograms in Figure 3 indicate, some twenty-first-century articles in *American Antiquity* and *Journal of Field Archaeology* do mention sampling or samples, and there has even been an encouraging “uptick” in the last few years, but rarely do they describe these explicitly as probability samples. In a disproportionate stratified random cluster sample of all research articles in *American Antiquity*, *Journal of*

Archaeological Science (JAS), *Journal of Field Archaeology* (JFA), and *Journal of Archaeological Method and Theory* (JAMT) from 2000 to 2019 (Supplemental Text 4), $24 \pm 1.1\%$ of articles mention some kind of sample without specifying what kind of sample it is. Furthermore, some that explicitly use probability samples after 2000 involve samples collected in the 1980s (e.g., Varien et al. 2007) rather than presenting any new sample. Few acknowledge use of convenience samples ($0.9 \pm 0.2\%$), but it seems likely that most of the unspecified samples were also of this type.

A few sampling-related articles in these and other journals show originality or new approaches (e.g., Arakawa et al. 2013; Burger et al. 2004; Perreault 2011; Prasciunas 2011). We also find random sampling in simulations (e.g., Deller et al. 2009). Despite a few bright spots, however, most of the articles in this period make no use of sampling theory, do not explicitly identify the population they sampled, and do not account for cluster sampling in their statistics—if they provide sampling errors at all. Some in my sample use “sampling” as a synonym for “collecting” ($1.3 \pm 0.25\%$ overall but almost 3% in *American Antiquity* and JFA) and “systematic sampling” in a nonstatistical sense ($0.4 \pm 0.1\%$ but almost 2% in JAMT), or they use “sample” as a synonym for “specimen” (e.g., individual lithics $3.0 \pm 0.5\%$, bones or bone fragments $10.7 \pm 0.8\%$, most of the latter in JAS).

Many authors who mention “samples” actually base analyses on all available evidence, such as all the pottery excavated at a site, or all known Clovis points from a region ($8.6 \pm 0.7\%$). These are only samples in the sense of convenience samples, and they are arguably populations in the present.

One of the most common practices is to use “sample” only in the sense of a small amount of material, such as some carbon for dating or a few milligrams of an artifact removed for archaeometric analysis (carbon samples $15.3 \pm 0.7\%$, pottery $5.3 \pm 0.74\%$, other $20.5 \pm 1.1\%$), selected without sampling theory. *American Antiquity* was most likely to refer to carbon specimens as samples. Many studies use “flotation sample” to refer to individual flotation volumes ($4.5 \pm 0.4\%$, mainly in *American*

Antiquity and *JFA*) rather than the entire sample from a site or context, and we see similar usage of “soil sample” even more often ($16 \pm 0.8\%$, most often in *JFA*).

Articles on regional survey after 2000 often mention sampling with no reference to sampling theory ($14.5 \pm 8.6\%$). Some claim “full-coverage” but employ systematic transect samples ($0.3 \pm 0.16\%$). Some articles not in this sample claim to use stratified sampling but actually selected “tracts” within strata purposively or by convenience (e.g., Tankosić and Chidiroglou 2010:13; Tartaron 2003:30). Many of these may have been effective in achieving their goals, but it is unclear whether stratification was effective or if they controlled biases in estimates. Among the encouraging exceptions, Parcero-Oubiña and colleagues (2017) use a stratified sample of agricultural plots in Chile, having estimated the sample size they would need in each stratum to achieve desired confidence intervals, and PPS random point sampling in each stratum to select plots.

In site excavation, purposive sampling is typical, while selection of excavated contexts for detailed analysis often occurs with little or no explanation (e.g., Douglass et al. 2008). Excavation directors understandably use expertise and experience or, at times, deposit models (sometimes based on purposive auger samples) to decide which parts of sites might best provide evidence relevant to their research questions (Carey et al. 2019). However, at least one study outside my sample used spatial sampling to estimate the number of features (Welch 2013).

Justifiably, purposive sampling dominates best practice in radiocarbon dating (Calabrisotto et al. 2017), whereas sampling for micromorphology, pollen, and plant macroremains tends to be systematic within vertical columns, and selection of column locations is purposive, if described at all (e.g., Pop et al. 2015). Alternatively, sampling protocols for plant remains may involve a type of cluster sample with a single, standardized sediment volume (e.g., 10 L) from every stratigraphic context or feature in an excavation (e.g., Gremillion et al. 2008). At least one case involves flotation of all contexts in their entirety (Mrozowski et al. 2008), not a sample at all. One article explicitly calculates

sample sizes needed at desired levels of error and confidence for comparing Korean assemblages of plant remains (Lee 2012), while Hiscock (2001) explicitly addresses sample size in artifact analyses.

In heritage management, our evidence comes more from regulatory frameworks and guidelines (Supplemental Text 5) than from the articles reviewed for Supplemental Texts 1 and 4. Although much of this work looks like sampling, the main purpose of regional CRM inventory surveys in most cases is to detect and document archaeological resources, not just sample them (e.g., MTCS 2011:74). Many North American guidelines specify systematic survey by pedestrian transects or shovel tests, but their purpose is primarily site discovery, not parameter estimation (see Shott 1985, 1987, 1989), and selection of in-site areas for excavation tends to be purposive (Neumann and Sanford 2010:174). Some jurisdictions do offer flexibility, however. The Bureau of Land Management (BLM) “does not discuss how to design class II surveys because the methods and theories of sampling are continually being refined” (BLM 2004:21B4). Meanwhile, Wisconsin’s standards explicitly address spatial probability sampling in research design (Kolb and Stevenson 1997:34). A memorandum of agreement among stakeholders in the Permian Basin has oil and gas developers pay into a pool that funds archaeological research and management in this part of New Mexico without tying it to project areas (Larralde et al. 2016; Schlanger et al. 2013). This approach allows more flexible research designs (Shott 1992), including, where warranted, probability sampling.

It may not seem obvious that sampling is relevant to experimental archaeology but, for almost a century, probability has had a role in ensuring that confounding factors, such as differential skill among flintknappers or variations in bone geometry, do not compromise the results of experiments (Fisher 1935). In a way, experimenters sample from all possible combinations of “treatments.” Yet, sampling theory has had little impact on experimental archaeology. In introducing a volume on this topic, Outram (2008) makes no mention of statistical sampling or randomization, nor do any of the articles in that volume. Some articles in Ferguson (2010) and Khreisheh

and colleagues (2013) discuss experimental controls or confounding factors, but none highlights randomization, arguably the most important protocol. Only Harry (2010:35) employs randomization but draws no attention to its importance. Of the articles in Supplemental Text 4, $8.6 \pm 0.7\%$ described experiments that made no use of randomization. Most of these were in *JAS*. Some encouraging exceptions highlight validity, confounding variables, and use of randomness (Lin et al. 2018 and articles cited there). Excluding randomness from experimental protocols risks confusing variables of interest with such variables as experimenter fatigue or the order in which a flintknapper selects cores (cf. Daniels 1978).

More generally, claims for random samples often have no supporting description (e.g., Benedict 2009:157). It is difficult to assess whether these were true probability samples or just haphazard (“grab-bag”) convenience samples; in Supplemental Text 1, I give them the benefit of the doubt. As noted, 24% of articles in Supplemental Text 4 use samples without stating their sampling methods ($31 \pm 5\%$ of articles in *American Antiquity*). Some researchers mix a random sample with a judgment sample without providing data that would allow us to disentangle them (e.g., Vaughn and Neff 2000).

Sometimes we find such puzzling claims as “although they were collected from a single unit . . . , these bones are a fairly representative sample of the faunal assemblage” (Flad 2005: 239) or “although $\frac{1}{8}$ -inch screens can cause significant biases . . . exceptional preservation . . . along with the dearth of very small fish . . . suggest that our samples are relatively representative” (Rick et al. 2001:599). One article admits that three houses constitute a small sample but claims it is a “reasonable representative sample” of some unidentified population (Hunter et al. 2014:716). Baseless assertions that samples are “representative” occur in $2.8 \pm 0.4\%$ of articles in Supplemental Text 4, but they are especially prevalent in *JFA* ($6.8 \pm 2.3\%$). Other authors assume that a large sample size is enough to make their samples representative. It is possible that some of these projects did employ probability sampling, but if so, they did not describe it (e.g., Spencer et al. 2008).

At least one study, apparently based on a convenience sample, claims that “relatively consistent” artifact densities across a site indicate that patterns identified “did not result from sampling bias” (Loendorf et al. 2013:272). Another suggests that a pattern it identifies “does not stem from vagaries in sampling” (Yasur-Landau et al. 2015:612) without describing any sampling design that would have assured this. Yet another asserts that “five . . . fragments selected nonrandomly and another five . . . indiscriminately . . . comprise a random sample” (Schneider 2015:519).

Other authors proudly ignore statisticians’ warning that “errors introduced by using simple-random-sampling formulas for . . . cluster samples can be extremely serious” (Blalock 1979:571). Campbell acknowledges use of a cluster sample but implies that the “statistically-informed” are being pedantic when he claims “conventional statistics . . . have been shown to work well” (Campbell 2017:15). Poteate and Fitzpatrick (2013) similarly use the statistics for simple random element sampling on simulated cluster samples, yielding incorrect confidence intervals on such measures as NISP. They also ignore that there is no reason to expect a small sample to yield the same MNI or taxonomic richness as a whole population, since these are very different levels of aggregation, and call to mind Hole’s (1980:226) disparagement of many such simulations.

These examples suggest a field that has mostly given up on sampling theory, notwithstanding Figure 3’s slight uptick in the last few years and the presence of some very good exceptions to the trend. Not all archaeological research should employ probability sampling, and we all make mistakes, but some statements like those just mentioned pose serious concerns. Too many articles mention samples with no indication of whether they were probabilistic or not but treat them as representative. Many use “sample” simply to refer to specimens, selections, or fragments, and “number of samples” to mean sample size. Finally, authors of some studies that did not employ probability sampling are not shy about blaming the failure of results to meet expectations on “probable sampling error” rather than on incorrect hypotheses or methods.

This poses grave challenges to the validity of the articles' conclusions.

Can We Revive Probability Sampling?

Not all of archaeology benefits from probability sampling. We are not always interested in the "typical" or "average," but rather in targeting significant anomalies, optimal preservation, or evidence relevant to specific hypotheses. In some contexts, however, inattention to sample quality has real consequences.

Sampling theory remains important whenever we want to generalize about or compare populations without observing them in their entirety, such as estimating total roofed area in an Iroquoian village without excavating an entire site (cf. Shott 1987). Inattention to sampling could lead to the erroneous inference of significant difference between sites, or change over time, when there is not—or, conversely, failure to identify significant differences or changes that actually did occur.

It also has a role in experimental archaeology. Experimenters must demonstrate that they are measuring what they purport to be measuring by controlling for confounding variables, such as skill differences among participants or quality differences in materials. One tool for this is randomization, much like using a probability sample from a population of potential experimental configurations.

So, how might we encourage more serious attention to and more widespread use of thoughtful sampling in archaeology?

In place of "cookbook" descriptions of sampling, textbooks could contextualize sampling within problem-oriented research design. They could encourage students to think about certain situations in which sampling would be helpful, and other situations in which more targeted research would be more useful. The key is to ensure validity of observations and conclusions (Daniels 1978).

Course curricula could include courses that prepare students to understand sampling as a practical aspect of research design, not just regional or site survey. Rather than teach textbook sample designs, we could encourage students to think critically about preventing their

own preconceptions or vagaries of research from yielding biased characterizations of sites, artifacts, or assemblages, or inferences of dubious validity.

We could also be more precise with terminology. Should we restrict the word "sample" to subsets of observations from a larger population? And should we replace "flotation sample," "NAA sample," and the like with "flotation volume," "NAA specimen," and so on? Even "carbon sample" deserves a better term to indicate that its selection has nothing to do with sampling theory. Oxymorons such as "sample population" and "sample parameter" have no place in our literature. We need to clarify whether a "stratified" sample is stratified in the statistical sense or just a specimen from a stratified deposit, and "systematic," in the context of sampling, should not just be a synonym for "methodical."

Conclusions

I suggest the following sampling "takeaways":

- (1) Probability sampling is not always appropriate but, when generalizing about or comparing populations on the basis of limited observations, failure to employ probability sampling may threaten the validity of results.
- (2) Sampling is not only for spatial situations, but also assemblages of artifacts, faunal and plant remains, temper or chemical evidence in pottery or lithics, and many kinds of experiments.
- (3) Cluster sampling, ubiquitous in archaeology, requires appropriate statistics for estimating variance. Ignoring this affects the outcomes of statistical tests.
- (4) Some archaeological samples are PPS samples, also requiring appropriate statistics to avoid bias.
- (5) Stratified sampling requires relevant prior information and follow-up evaluation to ensure that criteria for stratification were effective.
- (6) Straightforward methods are available to ensure that sample sizes are adequate; arbitrary sampling fractions are worthless.
- (7) When in doubt, talk to a statistician.

Sampling theory has had a rocky ride in archaeology. Negative perceptions of scientism, promotion of full-coverage survey, and flaws in past sampling-based research probably discouraged archaeologists' interest in formal sampling methods.

Yet the need for valid inferences persists, perhaps all the more as we increasingly mine "Big Data" from legacy projects. Probability sampling has the potential, in conjunction with well-conceived purposive selection, to contribute to archaeological research designs that are thoughtful, efficient, and able to yield valid inferences. We should not let misconceptions of the 1970s or 1990s deter us from taking full advantage of its well-established methods.

Acknowledgments. I am grateful to R. Lee Lyman and several anonymous reviewers for their insightful, constructive, and extremely useful comments on previous versions of this article. I would also like to thank Gary Lock, Piraye Hacıgüzeller, and Mark Gillings for the invitation that unexpectedly led me to write it, and Sophia Arts for editorial work and assistance in compiling statistics on published articles.

Data Availability Statement. No original data were used in this article.

Supplemental Materials. For supplemental material accompanying this article, visit <https://doi.org/10.1017/aaq.2020.39>.

Supplemental Text 1. List of publications in *American Antiquity* (1960–2019) and *Journal of Field Archaeology* (1974–2019) used for the histograms in Figure 3. Note that there was an interruption in *Journal of Field Archaeology* from 2002 until early 2004.

Supplemental Text 2. Distribution of sampling topics covered in a selection of introductory undergraduate texts, as well as some more specialized ones, over several decades. The list excludes texts that only cover culture history and, for texts with many editions, only includes one early and one recent edition.

Supplemental Text 3. The 25 highest-ranked international undergraduate programs in archaeology (Quacquarelli Symonds Limited 2019), listed alphabetically, and their 2018–2019 or 2019–2020 requirements relevant to sampling, according to program websites. Where universities had multiple archaeology programs, the table reflects the one related to anthropology or prehistory. Information that was not available publicly online is marked by (?). As no information on archaeological programs at the Sorbonne was available online, the sample size is $n = 24$.

Supplemental Text 4. Disproportionate stratified random cluster sample of the population of research articles and reports in *American Antiquity*, *Journal of Archaeological Science*, *Journal of Field Archaeology*, and *Journal of*

Archaeological Method and Theory from January 2000 to December 2019, summarizing the proportions of articles that use "sample" or sampling in various ways, along with evaluation of the stratification's effectiveness.

Supplemental Text 5. Examples of standards and guidelines for archaeological fieldwork in the heritage (CRM) industry.

References Cited

- Ambrose, Wal R.
1967 Archaeology and Shell Middens. *Archaeology and Physical Anthropology in Oceania* 2:169–187.
- Arakawa, Fumiyasu, Christopher Nicholson, and Jeff Rasic
2013 The Consequences of Social Processes: Aggregate Populations, Projectile Point Accumulation, and Subsistence Patterns in the American Southwest. *American Antiquity* 78:147–165.
- Banning, Edward B.
1996 Highlands and Lowlands: Problems and Survey Frameworks for Rural Archaeology in the Near East. *Bulletin of the American Schools of Oriental Research* 301:25–45.
- 2000 *The Archaeologist's Laboratory: The Analysis of Archaeological Data*. Kluwer Academic/Plenum Publishing, New York.
- 2002 *Archaeological Survey*. Kluwer Academic/Plenum, New York.
- 2020 Spatial Sampling. In *Archaeological Spatial Analysis: A Methodological Guide*, edited by Mark Gillings, Piraye Hacıgüzeller, and Gary Lock, pp. 41–59. Routledge, Abingdon, UK.
- Bayman, James M.
1996 Shell Ornament Consumption in a Classic Hohokam Platform Mound Community Center. *Journal of Field Archaeology* 23:403–420.
- Bender, Susan J.
2000 A Proposal to Guide Curricular Reform for the Twenty-First Century. In *Teaching Archaeology in the Twenty-first Century*, edited by Susan J. Bender and George S. Smith, pp. 31–48. SAA Press, Washington, DC.
- Bender, Susan J., and George S. Smith (editors)
2000 *Teaching Archaeology in the Twenty-first Century*. SAA Press, Washington, DC.
- Benedict, James B.
2009 A Review of Lichenometric Dating and Its Applications to Archaeology. *American Antiquity* 74:143–172.
- Berggren, Åsa, and Ian Hodder
2003 Social Practice, Method, and Some Problems of Field Archaeology. *American Antiquity* 68:421–434.
- Binford, Lewis R.
1964 A Consideration of Archaeological Research Design. *American Antiquity* 29:425–441.
- Blalock, Hubert M., Jr.
1979 *Social Statistics*. 2nd ed. McGraw-Hill, New York.
- BLM (Bureau of Land Management)
2004 8110 – Identifying and Evaluating Cultural Resources. United States Department of the Interior. Electronic document, https://www.blm.gov/sites/blm.gov/files/uploads/mediacenter_blmpolicymanual8110_0.pdf, accessed January 23, 2020.
- Buck, Caitlin E., William G. Cavanagh, and C. Litton
1996 *Bayesian Approach to Interpreting Archaeological Data*. John Wiley & Sons, London.

- Burger, Oskar, Lawrence C. Todd, Paul Burnett, Thomas J. Stohlgren, and Doug Stephens
2004 Multi-Scale and Nested-Intensity Sampling Techniques for Archaeological Survey. *Journal of Field Archaeology* 29:409–423.
- Calabrisotto, C. Scirè, Mariaelena Amadio, Mariaelena Fedi, Lucia Liccioli, and Luca Bombardieri
2017 Strategies for Sampling Difficult Archaeological Contexts and Improving the Quality of Radiocarbon Data: The Case of Erimi Laonin tou Porakou, Cyprus. *Radiocarbon* 59:1919–1930.
- Campbell, Sarah K.
1981 *The Duwamish No. 1 Site, a Lower Puget Sound Shell Midden*. Office of Public Archaeology Research Report 1. University of Washington, Seattle.
- Campbell, Greg
2017 “What Do I Do with All These Shells?” Basic Guidance for the Recovery, Processing, and Retention of Archaeological Marine Shells. *Quaternary International* 42:713–720.
- Carey, Chris, Andy J. Howard, Jane Corcoran, David Knight, and Jen Heathcote
2019 Deposit Modeling for Archaeological Projects: Methods, Practice, and Future Developments. *Geoarchaeology* 34:495–505.
- Cherry, John F., Clive Gamble, and Stephen Shennan (editors)
1978 *Sampling in Contemporary British Archaeology*. BAR British Series 50. British Archaeological Reports, Oxford.
- Collins-Elliott, Stephen A.
2017 Bayesian Inference with Monte Carlo Approximation: Measuring Regional Differentiation in Ceramic and Glass Vessel Assemblages in Republican Italy, ca. 200 BCE–20 CE. *Journal of Archaeological Science* 80:37–49.
- Cowgill, George L.
1964 The Selection of Samples from Large Sherd Collections. *American Antiquity* 29:467–473.
1970 Some Sampling and Reliability Problems in Archaeology. In *Archéologie et Calculateurs*, edited by Jean-Claude Gardin, pp. 161–175. CNRS, Paris.
1975 A Selection of Samplers: Comments on Archaeo-Statistics. In *Sampling in Archaeology*, edited by James Mueller, pp. 258–274. University of Arizona Press, Tucson.
1990 Toward Refining Concepts of Full-Coverage Survey. In *The Archaeology of Regions: The Case for Full-Coverage Survey*, edited by Suzanne K. Fish and Stephen A. Kowalewski, pp. 249–259. Smithsonian Institution, Washington, DC.
1993 Distinguished Lecture in Archaeology: Beyond Criticizing New Archaeology. *American Anthropologist* 95:551–573.
2015 Some Things I Hope You Will Find Useful Even If Statistics Isn’t Your Thing. *Annual Review of Anthropology* 44:1–14.
- Daniels, Stephen G. H.
1978 Implications of Error: Research Design and the Structure of Archaeology. *World Archaeology* 19: 29–35.
- Deller, D. Brian, Christopher J. Ellis, and James R. Keron
2009 Understanding Cache Variability: A Deliberately Burned Early Paleoindian Tool Assemblage from the Crowfield Site, Southwestern Ontario, Canada. *American Antiquity* 74:371–397.
- Douglass, Matthew J., Simon J. Holdaway, Patricia C. Fanning, and Justin I. Shiner
2008 An Assessment and Archaeological Application of Cortex Measurement in Lithic Assemblages. *American Antiquity* 73:513–526.
- Drennan, Robert D.
2010 *Statistics for Archaeologists: A Common Sense Approach*. Springer, New York.
- Dunnell, Robert C.
1984 The Ethics of Archaeological Significance Decisions. In *Ethics and Values in Archaeology*, edited by Ernestene L. Green, pp. 62–74. Free Press, New York.
- Ebert, James I.
1992 *Distributional Archaeology*. University of New Mexico Press, Albuquerque.
- Falconer, Steven E.
1995 Rural Responses to Early Urbanism: Bronze Age Household and Village Economy at Tell el-Hayyat, Jordan. *Journal of Field Archaeology* 22:399–419.
- Ferguson, Jeffry R. (editor)
2010 *Designing Experimental Research in Archaeology: Examining Technology through Production and Use*. University Press of Colorado, Boulder.
- Fish, Suzanne K., and Stephen A. Kowalewski (editors)
1990 *The Archaeology of Regions: The Case for Full-Coverage Survey*. Smithsonian Institution, Washington, DC.
- Fisher, Sir Ronald A.
1935 *The Design of Experiments*. Oliver and Boyd, Edinburgh.
- Flad, Rowan K.
2005 Evaluating Fish and Meat Salting at Prehistoric Zhongba, China. *Journal of Field Archaeology* 30:231–253.
- Flannery, Kent V.
1976 The Trouble with Regional Sampling. In *The Early Mesoamerican Village*, edited by Kent V. Flannery, pp. 159–160. Academic Press, New York.
- Florida Division of Historical Resources
2016 *Module Three, Guidelines for Use by Historic Preservation Professionals*. Division of Historical Resources, Florida Department of State, Tallahassee. Electronic document, <https://dos.myflorida.com/media/31394/module3.pdf>, accessed January 23, 2020.
- Given, Michael, A. Bernard Knapp, Nathan Meyer, Timothy E. Gregory, Vasiliki Kassianidou, Jay Stratton Noller, Lisa Wells, Neil Urwin, and Haddon Wright
1999 The Sydney Cyprus Survey Project: An Interdisciplinary Investigation of Long-Term Change in the North Central Troodos, Cyprus. *Journal of Field Archaeology* 26:19–39.
- Gremillion, Kristen, Jason Windingstad, and Sarah C. Sherwood
2008 Forest Opening, Habitat Use, and Food Production on the Cumberland Plateau, Kentucky: Adaptive Flexibility in Marginal Settings. *American Antiquity* 73:387–411.
- Haggett, Peter
1965 *Locational Analysis in Human Geography*. Edward Arnold, London.
- Harry, Karen G.
2010 Understanding Ceramic Manufacturing Technology: The Role of Experimental Archaeology. In *Designing Experimental Research in Archaeology*, edited by J. R. Ferguson, pp. 13–45. University Press of Colorado, Boulder.

- Hill, James N.
1970 *Broken K Pueblo: Prehistoric Social Organization in the American Southwest*. University of Arizona Press, Tucson.
- Hiscock, Peter
2001 Sizing Up Prehistory: Sample Size and Composition of Artefact Assemblages. *Australian Aboriginal Studies* 1:48–62.
- Hole, Bonnie L.
1980 Sampling in Archaeology: A Critique. *Annual Review of Anthropology* 9:217–234.
- Holtzman, Richard C.
1979 Maximum Likelihood Estimation of Fossil Assemblage Composition. *Paleobiology* 5:77–89.
- Hunter, Ryan, Stephen W. Silliman, and David B. Landon
2014 Shellfish Collection and Community Connections in Eighteenth-Century Native New England. *American Antiquity* 79:712–729.
- Johnson, Matthew
1999 *Archaeological Theory: An Introduction*. Blackwell, Oxford.
- Judge, W. James, James I. Ebert, and Robert K. Hitchcock
1975 Sampling in Regional Archaeological Survey. In *Sampling in Archaeology*, edited by James W. Mueller, pp. 82–123. University of Arizona Press, Tucson.
- Kamermans, Hans
1995 Survey Sampling, Right or Wrong. In *Computer Applications and Quantitative Methods in Archaeology 1994*, edited by Jeremy Huggett and Nick Ryan, pp. 123–126. BAR International Series 600. British Archaeological Reports, Oxford.
- Khreisheh, Nada N., Danielle Davies, and Bruce A. Bradley
2013 Extending Experimental Control: The Use of Porcelain in Flaked Stone Experimentation. *Advances in Archaeological Practice* 1:38–46.
- Kintigh, Keith W.
1990 Comments on the Case for Full-Coverage Survey. In *The Archaeology of Regions: The Case for Full-Coverage Survey*, edited by Suzanne K. Fish and Stephen A. Kowalewski, pp. 237–242. Smithsonian Institution, Washington, DC.
- Kolb, Jennifer L., and Katherine Stevenson (editors)
1997 Guidelines for Public Archaeology in Wisconsin. Wisconsin Archaeological Survey. Electronic document, http://www4.uwm.edu/Org/was/WASurvey/WAS_Guidlines_files/WAS_Guidelines_DOC.pdf, accessed January 23, 2020.
- Kowalewski, Stephen A.
1990 Merits of Full-Coverage Survey: Examples from the Valley of Oaxaca, Mexico. In *The Archaeology of Regions: A Case for Full-Coverage Survey*, edited by Suzanne K. Fish and Stephen A. Kowalewski, pp. 33–85. Smithsonian Institution, Washington, DC.
- Kuna, Martin
1998 Method of Surface Artefact Survey. In *Space in Prehistoric Bohemia*, edited by Evžen Neustupný pp. 77–83. Institute of Archaeology, Czech Academy of Sciences, Prague.
- Laralde, Signa, Martin Stein, and Sarah H. Schlanger
2016 The Permian Basin Programmatic Agreement after Seven Years of Implementation. *Advances in Archaeological Practice* 4:149–160.
- Lee, Gyoung-ah
2012 Taphonomy and Sample Size Estimation in Paleoethnobotany. *Journal of Archaeological Science* 39:648–655. DOI: [10.1016/j.jas.2011.10.025](https://doi.org/10.1016/j.jas.2011.10.025).
- Leonard, Robert D.
1987 Incremental Sampling in Artifact Analysis. *Journal of Field Archaeology* 14:498–500.
- Lin, Sam C., Zeljko Rezek, and Harold L. Dibble
2018 Experimental Design and Experimental Inference in Stone Artifact Archaeology. *Journal of Archaeological Method and Theory* 25:663–688.
- Loendorf, Chris R., Craig M. Fertelmes, and Barnaby V. Lewis
2013 Hohokam to Akimel O'odham: Obsidian Acquisition at the Historic Period Sacate Site (GR-909), Gila River Indian Community, Arizona. *American Antiquity* 78:266–284.
- Lovis, William A.
1976 Quarter Sections and Forests: An Example of Probability Sampling in the Northeastern Woodlands. *American Antiquity* 41:364–372.
- MacDonald, Burton, Lawrence A. Pavlish, and Edward B. Banning
1979 The Wādī al-Hasā Survey 1979: A Preliminary Report. *Annual of the Department of Antiquities of Jordan* 24:169–183.
- McAnany, Patricia A., and Sarah M. Rowe
2015 Re-Visiting the Field: Collaborative Archaeology as Paradigm Shift. *Journal of Field Archaeology* 40:499–507.
- McManamon, Francis P.
1981 Probability Sampling and Archaeological Survey in the Northeast: An Estimation Approach. In *Foundations of Northeast Archaeology*, edited by Dean R. Snow, pp. 195–227. Academic Press, New York.
- Meadow, Richard H.
1980 Animal Bones: Problems for the Archaeologist Together with Some Possible Solutions. *Paléorient* 6:65–77.
- Mrozowski, Stephen A., Maria Franklin, and Leslie Hunt
2008 Archaeobotanical Analysis and Interpretation of Enslaved Virginian Plant Use at Rich Neck Plantation (44WB52). *American Antiquity* 73:699–728.
- MTCS (Ministry of Tourism, Culture and Sport, Ontario)
2011 Standards and Guidelines for Consultant Archaeologists. Queen's Printer for Ontario, Toronto. Electronic document, http://www.mtc.gov.on.ca/en/publications/SG_2010.pdf, accessed January 23, 2020.
- Muckle, Robert J.
2014 *Introducing Archaeology*. 2nd ed. University of Toronto Press, Toronto.
- Mueller, James W.
1974 The Use of Sampling in Archaeological Survey. *Memoirs of the Society for American Archaeology* 28. *American Antiquity* 39(2, Pt. 2):1–91.
- Mueller, James W.
1975a Archaeological Research as Cluster Sampling. In *Sampling in Archaeology*, edited by James W. Mueller, pp. 33–41. University of Arizona Press, Tucson.
- Mueller, James W. (editor)
1975b *Sampling in Archaeology*. University of Arizona Press, Tucson.
- Nance, Jack D.
1981 Statistical Fact and Archaeological Faith: Two Models in Small-Sites Sampling. *Journal of Field Archaeology* 8:151–165.
- Nance, Jack D., and Bruce F. Ball
1986 No Surprises? The Reliability and Validity of Test Pit Sampling. *American Antiquity* 51:457–483.

- Neumann, Thomas W., and Robert M. Sanford
2010 *Practicing Archaeology: An Introduction to Cultural Resources Archaeology*, 2nd ed. AltaMira, Lanham, Maryland.
- O'Brien, Michael J., and Dennis E. Lewarch (editors)
1979 "Recent Approaches to Surface Data and Sampling." Special issue, *Western Canadian Journal of Anthropology* 8(3).
- Opitz, Rachel S., Krysta Ryzewski, John F. Cherry, and Brenna Moloney
2015 Using Airborne LIDAR Survey to Explore Historic-Era Archaeological Landscapes of Montserrat in the Eastern Caribbean. *Journal of Field Archaeology* 40:523–541.
- Orton, Clive
2000 *Sampling in Archaeology*. Cambridge University Press, Cambridge.
- Outram, Alan K.
2008 Introduction to Experimental Archaeology. *World Archaeology* 40:1–6.
- Parcero-Oubina, César, Pastor Fábrega-Álvarez, Diego Salazar, Andrés Troncoso, Frances Hayashida, Mariela Pino, César Borie, and Ester Echenique
2017 Ground to Air and Back Again: Archaeological Prospection to Characterize Prehispanic Agricultural Practices in the High-Altitude Atacama (Chile). *Quaternary International* 435 B:98–113.
- Parsons, Jeffrey R.
1990 Critical Reflections on a Decade of Full-Coverage Regional Survey in the Valley of Mexico. In *The Archaeology of Regions: The Case for Full-Coverage Survey*, edited by Suzanne K. Fish and Stephen A. Kowalewski, pp. 7–31. Smithsonian Institution, Washington, DC.
- Peacock, William R. B.
1978 Probabilistic Sampling in Shell Middens: A Case Study from Oronsay, Inner Hebrides. In *Sampling in Contemporary British Archaeology*, edited by John Cherry, Clive Gamble, and Stephen Shennan, pp. 177–190. British Archaeological Reports, Oxford.
- Perreault, Charles
2011 The Impact of Site Sample Size on the Reconstruction of Culture Histories. *American Antiquity* 76: 547–572.
- Phillips, Philip, James A. Ford, and James B. Griffin
1951 *Archaeological Survey in the Lower Mississippi Alluvial Valley, 1940–1947*. Papers of the Peabody Museum of Archaeology and Ethnology 25. Harvard University, Cambridge, Massachusetts.
- Plog, Stephen
1976 Relative Efficiencies of Sampling Techniques for Archaeological Surveys. In *The Early Mesoamerican Village*, edited by Kent V. Flannery, pp. 136–158. Academic Press, New York.
- Plog, Fred
1990 Some Thoughts on Full-Coverage Surveys. In *The Archaeology of Regions: The Case for Full-Coverage Survey*, edited by Suzanne K. Fish and Stephen A. Kowalewski, pp. 243–248. Smithsonian Institution, Washington, DC.
- Pop, Eduard, Corrie Bakels, Wim Kuijper, Herman Múcher, and Madeleine van Dijk
2015 The Dynamics of Small Postglacial Lake Basins and the Nature of Their Archaeological Record: A Case Study of the Middle Palaeolithic Site Neumark-Nord 2, Germany. *Geoarchaeology* 30:393–413.
- Poteate, Aaron S., and Scott M. Fitzpatrick
2013 Testing the Efficacy and Reliability of Common Zooarchaeological Sampling Strategies: A Case Study from the Caribbean. *Journal of Archaeological Science* 40:3693–3705.
- Prasciunas, Mary M.
2011 Mapping Clovis: Projectile Points, Behavior, and Bias. *American Antiquity* 76:107–126.
- Quacquarelli Symonds Limited
2019 QS World University Rankings, Archaeology. Electronic document, <https://www.topuniversities.com/university-rankings/university-subject-rankings/2019/archaeology>, accessed May 25, 2019.
- Ragir, Sonia
1967 A Review of Techniques for Archaeological Sampling. In *A Guide to Field Methods in Archaeology: Approaches to the Anthropology of the Dead*, edited by Robert F. Heizer, and John A. Graham, pp. 181–197. National Press, Palo Alto, California.
- Redman, Charles L.
1974 *Archaeological Sampling Strategies*. Addison-Wesley Module in Anthropology 55. Addison-Wesley, Reading, Massachusetts.
- Redman, Charles L., and Patty Jo Watson
1970 Systematic Intensive Surface Collection. *American Antiquity* 35:279–291.
- Renfrew, Colin, and Paul Bahn
2008 *Archaeology: Theories, Methods, and Practice*. 5th ed. Thames & Hudson, London.
- Richardson, Mary, and Byron Gajewski
2002 Archaeological Sampling Strategies. *Journal of Statistics Education* 10. DOI:10.1080/10691898.2003.11910693.
- Rick, Torben C., Jon M. Erlandson, and René L. Vellanoweth
2001 Paleocoastal Marine Fishing on the Pacific Coast of the Americas: Perspectives from Daisy Cave, California. *American Antiquity* 66:595–613.
- Rootenber, Sheldon
1964 Archaeological Field Sampling. *American Antiquity* 30:181–188.
- Sanders, William T., Jeffrey R. Parsons, and Robert S. Santley
1979 *The Basin of Mexico: Ecological Processes in the Evolution of a Civilization*. Academic Press, New York.
- Scheps, Sheldon
1982 Statistical Blight. *American Antiquity* 47:836–851.
- Schlanger, Sarah, George MacDonell, Signa Larralde, and Martin Sheen
2013 Going Big: The Permian Basin Memorandum of Agreement as a Fundamental Shift in Section 106 Compliance. *Advances in Archaeological Practice* 1:13–23.
- Schneider, Tsim D.
2015 Envisioning Colonial Landscapes Using Mission Registers, Radiocarbon, and Stable Isotopes: An Experimental Approach from San Francisco Bay. *American Antiquity* 80:511–529.
- Schuldenrein, Joseph, and Jeffrey H. Altschul
2000 Archaeological Education and Private-Sector Employment. In *Teaching Archaeology in the Twenty-first Century*, edited by Susan J. Bender and George S. Smith, pp. 59–64. SAA Press, Washington, DC.
- Shanks, Michael
1999 *Art and the Early Greek State: An Interpretive Archaeology*. Cambridge University Press, Cambridge.

- Shanks, Michael, and Christopher Tilley
1987 *Re-Constructing Archaeology: Theory and Practice*. Cambridge University Press, Cambridge.
- Shott, Michael L.
1985 Shovel-Test Sampling as a Site Discovery Technique: A Case Study from Michigan. *Journal of Field Archaeology* 12:457–468.
1987 Feature Discovery and the Sampling Requirements of Archaeological Evaluations. *Journal of Field Archaeology* 14:359–371.
1989 Shovel-Test Sampling in Archaeological Survey: Comments on Nance and Ball, and Lightfoot. *American Antiquity* 54:396–404.
1992 Commerce or Service? Models of Practice in Archaeology. In *Quandaries and Quests: Visions of Archaeology's Future*, edited by LuAnn Wandsnider, pp. 9–24. Center for Archaeological Investigations Occasional Paper 20. Southern Illinois University Press, Carbondale.
- Sørensen, Tim Flohr
2017 The Two Cultures and a World Apart: Archaeology and Science at a New Crossroads. *Norwegian Archaeological Review* 50:101–115.
- Spencer, Charles S., Elsa M. Redmond, and Christina M. Elson
2008 Ceramic Microtypology and the Territorial Expansion of the Early Monte Albán State in Oaxaca, Mexico. *Journal of Field Archaeology* 33:321–341.
- Stanford University
2019 Major, Bachelor of Arts in Archaeology. Stanford Archaeology Center, School of Humanities and Sciences. Electronic document, <https://archaeology.stanford.edu/academicsundergraduate-program/major>, accessed May 25, 2019.
- Sundstrom, Linea
1993 A Simple Mathematical Procedure for Estimating the Adequacy of Site Survey Strategies. *Journal of Field Archaeology* 20:91–96.
- Tankosić, Zarko, and Maria Chidiroglou
2010 The Karystian Kampos Survey Project: Methods and Preliminary Results. *Mediterranean Archaeology and Archaeometry* 10(3):11–17.
- Tartaron, Thomas F.
2003 The Archaeological Survey: Sampling Strategies and Field Methods. In *Landscape Archaeology in Southern Epirus, Greece*, edited by James Wiseman and Konstantinos Zachos, pp. 23–45. *Hesperia Supplements* 32. American School of Classical Studies at Athens, Athens.
- Thomas, David H.
1978 The Awful Truth about Statistics in Archaeology. *American Antiquity* 43:231–244.
1999 *Archaeology: Down to Earth*. Nelson Thomson Learning, Toronto.
- Thompson, Steven K., and George Seber
1996 *Adaptive Sampling*. John Wiley & Sons, New York.
- Ullah, Isaac, Paul Duffy, and Edward B. Banning
2015 Modernizing Spatial Micro-Refuse Analysis: New Methods for Collecting, Analyzing, and Interpreting the Spatial Patterning of Micro-Refuse from House-Floor Contexts. *Journal of Archaeological Method and Theory* 22:1238–1262. DOI:[10.1007/s10816-014-9223-x](https://doi.org/10.1007/s10816-014-9223-x).
- VanPool, Christine S., and Todd L. VanPool
1999 The Scientific Nature of Postprocessualism. *American Antiquity* 64:33–53.
- Varién, Mark D., Scott G. Ortman, Timothy A. Kohler, Donna M. Glowacki, and C. David Johnson
2007 Historical Ecology in the Mesa Verde Region: Results from the Village Ecodynamics Project. *American Antiquity* 72:273–299.
- Vaughn, Kevin J., and Hector Neff
2000 Moving beyond Iconography: Neutron Activation Analysis of Ceramics from Marcaya, Peru, an Early Nasca Domestic Site. *Journal of Field Archaeology* 27:75–90.
- Vesceius, Gary S.
1960 Archaeological Sampling: A Problem in Statistical Inference. In *Essays in the Science of Culture, in Honor of Leslie A. White*, edited by Gertrude E. Dole and Robert L. Carneiro, pp. 457–470. Crowell, New York.
- Wallace-Hadrill, Andrew
1990 The Social Spread of Roman Luxury: Sampling Pompeii and Herculaneum. *Papers of the British School at Rome* 58:145–192.
- Walsh, Michael R.
1998 Lines in the Sand: Competition and Stone Selection on the Pajarito Plateau, New Mexico. *American Antiquity* 63:573–593.
- Watson, Richard A.
1990 Ozymandias, King of Kings: Postprocessual Radical Archaeology as Critique. *American Antiquity* 55:673–689.
- Watson, Patty Jo, Stephen LeBlanc, and Charles L. Redman
1971 *Explanation in Archeology: An Explicitly Scientific Approach*. Columbia University Press, New York.
- Welch, Paul
2013 Designing a Sample of Cores to Estimate the Number of Features at a Site. *Advances in Archaeological Practice* 1:47–58.
- Williams, Leonard, David H. Thomas, and Robert Bettinger
1973 Notions to Numbers: Great Basin Settlements as Polythetic Sets. In *Research and Theory in Current Archaeology*, edited by Charles L. Redman, pp. 215–237. John Wiley & Sons, New York.
- Wobst, Martin H.
1983 We Can't See the Forest for the Trees: Sampling and the Shapes of Archaeological Distributions. In *Archaeological Hammers and Theories*, edited by James A. Moore and Arthur S. Keene, pp. 37–85. Academic Press, New York.
- Yasur-Landau, Assaf, Eric H. Cline, Andrew J. Koh, David Ben-Shlomo, Nimrod Marom, Alexandra Ratzlaff, and Inbal Samet
2015 Rethinking Canaanite Palaces? The Palatial Economy of Tel Kabri during the Middle Bronze Age. *Journal of Field Archaeology* 40:607–625.

Submitted July 29, 2019; Revised April 13, 2020; Accepted April 15, 2020