A DEFENSE OF SHOVEL-TEST SAMPLING: A REPLY TO SHOTT

Kent G. Lightfoot

Shott raises pertinent issues on the limitations of shovel-test sampling. I agree that alternative methods with more efficient discovery rates should be developed and tested. I argue, however, that shovel testing is the most efficient discovery technique now available for detecting buried cultural remains on a regional scale. In survey contexts where the probability of buried remains is high, shovel testing should remain a primary option.

Shott’s thoughtful comments are a welcome contribution to the growing literature on shovel-test sampling. He raises pertinent points on the difficulties of discovering cultural remains in wooded environments and/or buried contexts. Since shovel-test surveys sample very small fractions of space at very high labor costs, Shott argues that the method is inefficient and not cost-effective. He suggests that alternative survey methods should be employed in wooded environments.

I fully agree with Shott that alternative methods should be developed and tested for different survey conditions. Shovel testing is not a panacea for wooded environments. Other survey methods with better discovery rates should be explored as options.

I strongly disagree, however, with Shott’s (1989:396) characterization that shovel testing “is a survey method whose time, hopefully, has come and gone.” On the contrary, I argue that shovel testing is a very viable method for a specific kind of survey condition. It is the most effective method currently available for discovering buried cultural remains on a regional scale. I advocate it as a discovery technique when the probability is high that material remains are buried below ground surface to a depth of about 50 cm. Other discovery methods, such as remote sensing and chemical testing, which can be done in conjunction with shovel testing, are still not refined enough for detecting buried remains on a regional scale (see McManamon 1984:247–253). McManamon’s findings are supported by experimental use of soil resistivity, proton magnetometer and chemical-testing surveys over the last five years on Long Island, New York.

A well-designed, systematically executed shovel-test survey can produce realistic parameter estimates on the regional distribution of archaeological remains. That was the major point of my Shelter Island case example (Lightfoot 1986). One must recognize that shovel testing is a discovery technique predicated on regional sampling. It is unreasonable to expect that all archaeological manifestations will be detected. There is no survey method that I am aware of that will discover the total population of buried remains in a region. However, shovel-test surveys can generate estimates on the probability of discovering diverse-sized manifestations characterized by variable artifact densities and internal artifact distributions. I think future work, like that of Nance and Ball (1986), will continue to refine the discovery probabilities of shovel testing for different characteristics of the archaeological record.

With this discussion as background, I will address four points raised by Shott.

1) Site Definition. Shott argues that my positive appraisal of shovel-test sampling at Shelter Island is colored, in part, by problems of site definition. He notes (Shott 1989:400) that “By treating as separate sites artifact distributions that probably are sections of larger ones, Lightfoot underestimates site size.” The site definition employed at Shelter Island is strictly heuristic: A site is an archaeological manifestation characterized by clusters of artifacts/features distributed across the landscape. Shovel probes were excavated at 10-m intervals across survey sample units. When
“positive” probes (those containing cultural material) were discovered, additional shovel probes were excavated from that point every 2 m in the four cardinal directions (the “iron cross” technique, Lightfoot 1986:495). Using this technique, isolated finds were distinguished from broader manifestations containing multiple artifacts distributed over multiple shovel probes. The spatial distribution of cultural remains detected by the iron cross procedure was then cross-checked by resuming the excavation of shovel probes at 10-m intervals across the area.

The discussion on site definition is important for three reasons.

First, Shott’s (1989:399) statement that I treat “as distinct sites any two productive units . . . separated by an empty one” is misleading. Sites were defined by examining the broad spatial patterning of 10-m-interval shovel probes and multiple iron crosses. Positive and empty (negative) shovel probes were plotted on graph paper after the completion of survey sample units. Site boundaries were then defined by fall-off patterns of artifacts. Contrary to Shott’s discussion, many of the Shelter Island sites contain empty spaces. In the final report on the Shelter Island research, different kinds of archaeological manifestations were defined by artifact densities, percentages of positive probes that intersected site areas, and overall sizes of artifact scatters (see Lightfoot et al. 1987:40–43, 50, 59). More than 65 percent of the probes excavated in some sites failed to yield artifacts. (I should also add that Shott’s criticism of my method for calculating artifact densities is not applicable, since I did not generate them in the American Antiquity article. The artifact densities calculated in the final report take into account the area occupied by shovel probes within sites [see Lightfoot et al. 1987: 41–42].)

Second, despite our quibbling over site definition, it is clear that shovel testing can discover not only spatially extensive sites with high artifact densities, but many other kinds of archaeological manifestations. At Shelter Island we defined, in addition to the above, areally limited sites with high artifact densities, spatially extensive artifact scatters with low artifact densities, and isolated finds (see Lightfoot et al. 1987:50, 59). One should recognize that a smaller percentage of the total population of these latter manifestations actually will be found, but discovery probabilities can be generated to estimate their population parameters across the study area. For example, one can compare the projected densities of different kinds of manifestations across different habitat types, information that is very useful for many cultural resource management studies.

Third, there are no standardized, uniform criteria for defining sites of which I am aware. Nor is there one correct way of classifying them. Sites are defined in the field to classify specific kinds of archaeological manifestations that relate to the investigator’s research problem. This makes cross comparison of site densities difficult. However, Shott’s (1989:400) claim that my method subdivides “what most archaeologists would consider discrete sites into a number of smaller ones” and “artificially” inflates site densities is inaccurate. He has missed a very important point about the nature of my article. It compares the results of the Shelter Island survey to one that I directed in the American Southwest (Pinedale survey). In both surveys I used the same basic criteria (artifact fall-off patterns) to define sites.

(2) Comparable Results. This last point raises the issue of whether it is valid to compare the results of fieldwork undertaken by different survey methods. Shott (1989:400) argues that it is inappropriate to compare the results of shovel testing to surface survey, or for that matter any survey method that is not “sampled identically.” Since he is advocating the use of alternative survey methods, does this mean that each will be incomparable with another? While Shott is correct that difficulties currently exist in comparing the results of shovel testing and surface surveys, it is a far cry from being invalid. One does not have to employ identical sampling methods to estimate population parameters. The real issue is how well a researcher’s survey method estimates the actual population characteristics of a region. Survey work based on different sampling methods can be compared, as long as each provides a representative picture of the region. It is true that shovel testing needs further refinement. However, I am confident that as the discovery probabilities of shovel testing are clarified, so will our regional parameter estimates.

(3) Surface Survey. Like many other archaeologists, Shott (1989:400) characterizes surface survey as more effective than shovel testing. This is true as long as cultural remains are visible on the ground surface. The shortcoming of surface survey is that it only samples on a two-dimensional
plane, while archaeological manifestations exist in three dimensions. In regions where surface survey commonly is used (i.e., the American Southwest), archaeologists circumvent this problem by assuming that most or all cultural remains will be visible on the ground surface. However, in reality most large-scale survey projects in the American Southwest crosscut a diverse range of plant communities, soil types, topographic features, and drainage systems. Ground visibility and the potential for buried remains vary accordingly. Yet southwestern archaeologists rarely consider the probability that material remains will be buried, or that these probabilities will vary across different habitats. In truth, parameter estimates for surface surveys need to take into account the probability of not discovering buried remains. Otherwise surface surveys risk the problem of discovery bias, in which the most visible surface cultural remains are overrepresented at the expense of buried ones.

In comparing the efficiency of shovel testing with surface survey, we cannot ignore the problems of the latter. My fieldwork in the Southwest, the Northeast, and northern California suggests that buried remains may be more persistent than most North American archaeologists wish to acknowledge. Unless one can demonstrate that the target population of cultural remains is detectable readily on the ground surface (an assumption often made but rarely demonstrated), then survey methods should sample three-dimensional space. That is, before one assumes that surface surveys provide “remarkably thorough coverage” (Shott 1989:401), the probability that buried remains exist must be evaluated. The strength of shovel testing is that it tests for buried cultural remains while still providing thorough coverage of the ground surface. It combines surface survey with a baseline for evaluating the probability of buried remains across different plant communities, soil types, and topographic features.

(4) Implications. My major concern with Shott’s comments is the impact it may have on cultural resource management studies. Given the considerable labor investment that shovel testing requires, some archaeologists, government agencies, and developers may greet Shott’s argument as justification to terminate altogether subsurface testing. This would be a serious mistake, especially in regions where the probability of buried cultural remains is high. Until alternative methods are developed with more efficient discovery rates, shovel testing should remain a primary subsurface discovery technique.

The alternative methods suggested by Shott may have relatively limited application. The surface survey of clear-cut forest tracts should be explored as a viable option in some cases. However, the target population of cultural remains must be visible on the ground surface and evaluations of the overall cost-effectiveness of the method must consider the time and energy involved in clear cutting. In reality, most government agencies and developers will be rather skeptical of clear cutting forests, especially if they perceive that the method of evaluating environmental impact will be more adverse than the actual construction project. Another method discussed by Shott (1989:402) is monitoring “construction activities to exploit the ground-surface exposure they furnish.” Monitoring is a poor substitute for regional systematic survey. It severely limits management planning prior to construction, and it provides no real basis for estimating population parameters or predicting site locations. Furthermore, my experience with monitoring projects illustrates that once development begins and construction deadlines are set, developers are much less flexible in working with archaeologists.

Before dismissing shovel-test sampling in regions with high probabilities of buried remains, one must consider carefully the effectiveness of alternative methods. This cautionary statement is based on my own experience with cultural resource management work in Brookhaven Township, eastern Long Island, New York. Until recently, systematic shovel testing was not conducted as part of cultural resource surveys. Rather, surface surveys were employed with special consideration taken for walking agricultural fields, dirt roads, and erosional cuts. Without exception, these surveys tended to discover relatively few sites.

Recently I was requested by the Division of Environmental Protection, Brookhaven Township, to resurvey two proposed housing developments. The first survey area, a 68-ha tract in central Brookhaven, consisted of mixed deciduous woodlands and agricultural fields that had remained fallow for several years. A previous surface survey, where investigators walked road cuts and fields, yielded few cultural remains. In contrast, our excavation of 1,815 shovel probes at 10-m intervals across 16 survey sample units discovered a broad range of Euroamerican artifacts and some Native
American remains. The vast majority were not detected on the ground surface but were buried in the 30-cm plow zone. Contour-interval maps of artifact densities plotted by SYMAP exhibited a peak and fall-off pattern of unusually high numbers of historical-period materials. This peak corresponded to the location of the John Petty house, the first homestead in the survey area as identified by the 1838 U.S. Coastal Survey Map. While the original structure was destroyed as part of the continual modification of the existing agricultural fields, the spatial distribution of buried artifacts still identified the historically significant site (see Lightfoot et al. 1987).

The second survey of a 105-ha area was one of the last undeveloped parcels on Brookhaven's south shore. A previous surface survey of dirt roads and agricultural fields produced few cultural remains. In comparison, the subsurface survey of 1,860 shovel probes in 35 survey sample units yielded 9 sites, 4 of which were broad scatters of Native American artifacts (see Lightfoot et al. 1987). Since these remains were buried from 10 to 50 cm below surface, it is not surprising that they were missed in the previous survey. Furthermore, dirt roads where subsurface deposits could be inspected missed the sites, and the fallow agricultural fields produced relatively few artifacts on the surface.

The Long Island survey results contradict Shott's (1989:403) statement that shovel testing sanctions the "omission and unwitting destruction of many archaeological sites." On the contrary, since undertaking systematic shovel testing in Brookhaven Township, the number of sites recorded has increased dramatically, and a better understanding of the spatial distribution of both prehistoric and historical-period sites has resulted. I maintain that previous survey methods greatly underestimated the magnitude and extent of Native American sites in eastern Long Island (see Lightfoot et al. 1985). While shovel testing is much more labor intensive than other alternatives used to date, it has proven to be much more effective in discovering buried cultural remains.

In conclusion, I agree with Shott's basic point that more efficient survey methods need to be developed. While an advocate of shovel testing, I do not particularly enjoy digging multiple shovel probes in dense undergrowth. Given a viable alternative for sampling three dimensional space, I will gladly lay down my shovel and portable screen. However, despite experimentation with other methods, I know of no more effective survey method for discovering buried cultural remains on a regional scale. Given the high probability of buried remains in many regions of North America, it should remain a significant option of archaeologists in the near future.

REFERENCES CITED

Lightfoot, K.

Lightfoot, K., S. Gryzbowski, and F. Turano

Lightfoot, K., R. Kalin, O. Lindauer, and L. Wicks

Lightfoot, K., R. Kalin, and J. Moore

Lightfoot, K., J. Moore, S. Gryzbowski, F. Turano, R. Jewett, J. Muennig, and W. Rosfeld

McManamon, F. P.

Nance, J. D., and B. F. Ball

Shott, M. J.