American Antiquity, Jan 2003 v68 i1 p161-167


Author's Abstract: COPYRIGHT 2003 Society for American Archaeology

Owsley and Jantz (2001) allege that we (Swedlund and Anderson 1999) misunderstand and misrepresent the events surrounding the controversy over Kennewick Man, and that we misconstrue statements they have made regarding Spirit Cave Man. They then move to their own analysis of Gordon Creek Woman to demonstrate the value of their morphometric techniques in addressing questions of biological affinity. In this reply we clarify and amplify our position on the key issues on which they challenge us, and we evaluate their morphometric analysis of Gordon Creek Woman. To our previous call for bioarchaeologists to more explicitly acknowledge the political environment in which questions of biological affinity currently arise, we add our concern that the methodologies used have their own set of problems and limitations.

Full Text: COPYRIGHT 2003 Society for American Archaeology

We are pleased that our Forum article (1999) has provided a platform for other points of view, and we have been gratified by many of the responses we have received. Owsley and Jantz's (2001) published response to our piece covers a lot of ground, and while we would like to address each point we are aware of the very lengthy reply this would require. Therefore, in the interests of space and clarity, we will address only the most important and substantive issues. Among these, they allege that we lack understanding of the methods used for ascertaining the status of prehistoric remains. They also allege that we misunderstand the events surrounding the lawsuit filed by the eight scientists against the Army Corps of Engineers regarding their decision to repatriate the Kennewick remains to the Confederated Tribes of the Umatilla.

It is our opinion that once one accounts for the differing topics and the timing of appearance of our two Forum comments in American Antiquity, many of the issues they raise become irrelevant to our original piece. (1) Our purpose here is to focus on those that remain and to clarify our position on a few key points, but we wish to avoid participation in a heated exchange of opinions and accusations. Also, it seems that our prior conclusion regarding the desirable kinds of research design is consistent with the efforts of Owsley and Jantz, as our call for "sophisticated and nuanced methodologies" (p. 574) was a reference to (but not exclusively) the kinds of statistical approaches advocated by Owsley and Jantz, Steele and Powell (e.g., 1994) and others. Having said this, we need to acknowledge some reservations we have as to how these methods are being used in some cases.

The Science of Morphometric Comparison

Owsley and Jantz's analysis of Gordon Creek Woman provides an invitation for us to comment on a topic that was well outside the scope of our initial Forum article. We can appreciate the difficulty with which they undertook this evaluation, with only six cranial measurements deemed
acceptable for analysis. As they will have noted in the 1971 article on Gordon Creek, "the skull is represented by 45 fragments of which 39 were identifiable for use in the reconstruction.... A slight degree of post-mortem deformation is visually detectable in the right temporo-parietal area.... Reconstruction required several attempts to achieve satisfactory results.... Several small wooden supports were necessary.... Although post-mortem deformation has probably caused minor distortions, we believe that the reconstruction is reasonably accurate" (Breternitz et al. 1971:173). For analyses that depend on accuracy to the millimeter or even less, we believed (and still do) this was not a great candidate for extensive measurement and craniometric comparison, but we are gratified—if not concerned—that Owsley and Jantz, having read this, proceeded to trust the osteometrist (Swedlund) sufficiently to proceed. (2) In addition, Gordon Creek woman was slight and very short, with a stature estimated at between 4' 10" to 5' 2", given the regression models available at the time of publication. Skull measurements consequently are also on the low side of most distributions and osteologists are well aware of the overall effects of size on morphometric relationships, causing problems even when size is controlled.

As we further stated, given the very few remains that were candidates for Paleoindian status at that time, and given the condition of the skull, we did not consider comparisons meritorious beyond the general proportions expressed in the cephalic index. Owsley and Jantz and other investigators now regard the number of early remains sufficient to undertake this type of study. What we were hoping to show, then as now, is that there are other kinds of morphometric and mortuary information that made this discovery significant, regardless of considerations of biological affinity. Contrary to what Owsley and Jantz regard as our complaint about lack of citations to Gordon Creek Woman (e.g., p. 566), we were, in part, lamenting the lack of interest in these other aspects of early American remains and how Gordon Creek was now being included as a possible "Caucasoid" example (Anderson et al. 1997 and citations therein). We readily acknowledge that Owsley and Jantz did not use that term.

This brings us to the analysis that Owsley and Jantz provide in their Forum comment. From the 4,500 individuals of historical and prehistoric provenience at their disposal (p. 566), they select a total of 904 divided into seven samples ranging from Europe to Africa to the American Plains and Great Basin (see also 2001b:6). They then undertake the analysis using the six measurements they accept from Gordon Creek woman. Their conclusion, consistent with Steele and Powell (1994), is that Gordon Creek fits the pattern established for other ancient crania, namely, it plots similarly to European and Southern Asian populations. Then, from their own analysis they argue that Gordon Creek's greatest distances are from the two Native American samples selected and that it is on the extreme margins of both groups (p. 567).

While it is not possible here to explore in great detail this analysis, a few caveats are appropriate, notwithstanding our own reservations about the validity—to the millimeter—of our own measurements, given the fragmentary and distorted post-mortem condition of the skull. Our concerns have to do with the specimens and samples utilized, the spatial and temporal dimensions from which evidence is derived, the determinants of craniometric variability, and the specific statistical interpretations made.
Specimens, Samples, and Populations

Ascertaining the morphological status of Paleoindian remains in the procedure of Owsley and Jantz involves comparing the measurements of single specimen(s) from the early period to samples/populations of crania that date from a variety of time frames and locations, but preferably from the late precontact or early historical period in North America. We are all indebted to Howells (e.g., 1973, 1989), Owsley and Jantz, and others who have established these databases that permit comparisons to be made. However, we must carefully consider each database and sample and be attentive to sampling issues raised by the availability of remains and by those originally involved in assembling the databases. With the Howells database the crania for each sample generally come from sites in close proximity and time period, and number approximately 50 skulls for each sex (Howells 1973:6-7). Yet, the Great Basin sample Owsley and Jantz have generated and selected for comparison to Gordon Creek appears to comprise only 26 crania from several temporally unspecified sites (Jantz and Owsley 1997) and, therefore, is at best a very limited sample of a "population." (3)

To us there is a profound difference between saying that the crania(um) being compared are not like "a modern Indian" (quote attributed to Owsley in Morell 1998:192), or "fall outside the range of American Indian populations" (Owsley and Jantz 2001:567) and saying that the crania(um) being compared did not fit comfortably near the centroid of those samples with which it was being compared. It is also true of these sample populations that we cannot know the representativeness or randomness of the samples. We know that they are not random with respect to sex and age, and in some cases Howells is forthright in explaining that the museum curator and/or his hosts (1973:6,viii) selected from among the available skulls those thought to be most appropriate, or that Howells himself selected some crania because they were distinctive. They may be statistically representative, but we know that taphonomy, collection, curation, and subjective judgment all go into the selection of many of these samples.

Time, Space, and Form

The way in which Owsley and Jantz use the Mahalanobis $[D_{sup.2}]$ statistic is to compare a specimen to the database samples with regard to a suite (subset) of cranial measurements from the 65 they recommend, and to then place the specimen in some degree of probability of relationship to the comparison samples. The "distance" that Mahalanobis's $[D_{sup.2}]$ measures is a morphological, or measurement distance. Owsley and Jantz translate this into a two- or three-dimensional, principal components space. Their assumption is that a specimen being compared should plot closest to those population samples with which it is geographically most closely associated. If a specimen plots more closely to a population at a greater distance, or even on another continent, then it is assumed to be morphologically/genetically closer to that population sample. This is problematic, because a specimen that might date approximately 8,000 years older than its closest reference sample is not only separated by geographic distance but also by considerable temporal distance. We can translate this into very approximate generation times (e.g., 8,000 years/20) and quickly discover that we are talking about a "distance" of approximately 400 generations in which gene flow, drift, mutation, and natural selection have had an opportunity to operate between the specimen and its referents. (4) Add to this environmental plasticity and it is not at all surprising to us that some early Archaic American
specimens might plot more closely to Asian, Eurasian, and even European samples. With the exception of Powell and Neves's important work (1999), very few of the studies to date attempt to intercollate the so-called fossil crania and the modern crania with intermediates, such as middle or late Archaic samples. This would more accurately assess the possibility of continuities across long time "distances." In the one case we know of involving Kennewick, Powell and Rose (1999) find that Kennewick Man fits comfortably into the Indian Knoll skeletal sample dating from the middle-to-late Archaic period. Such tests go a long way in helping us understand trends in variability in the prehistoric occupants of North America.

Genes, Environment, and Skeletal Plasticity

Every osteologist teaching an introductory course in skeletal biology reminds his/her students that they should not think of the skeleton as the dry, brittle, hard bones that they see on the table before them, but rather as a dynamic and highly variable living tissue that responds to a wide array of genetic, environmental, dietary, life course, and even cultural forces to which it is exposed. Owsley and Jantz and others frequently make the comment that the face and general cranium is under strong genetic control (e.g., Jantz and Owsley 1997; Owsley and Jantz 2001:568). It is a mantra that, if repeated often enough, can make us complacent. Probably it is true, but only up to a point. Certainly the differences between rhesus monkeys and Homo sapiens must have a strong genetic determination, even though we have had trouble identifying specific differences in the genomes up to this point. However, although within-species differences can be expected to occur in part due to genetics, we have very little research and data verifying the magnitude and the genetic mechanisms involved.

By contrast, we have volumes of data on how the cranium responds to nutritional, dietary (they are not the same), and environmental forces within the life span, particularly during growth and development. This surely accounts for many examples of why crania show large variation within populations and differ through time with no apparent population replacement or significant gene flow. The classic in this regard was Boas's (1912) study of European immigrants to the U.S., and there are many studies of how the transition to agriculture in populations as diverse as North American Indians and Sudanese Nubians caused significant changes in facial morphology without evidence for population replacement (see Larsen 1997:226-236 for an excellent review). As Larsen notes, some of the most dramatic changes in skull size and shape have taken place in the transition from a hunting-gathering way of life to a sedentary agricultural one. Lifestyle, nutrition, and diet, and mechanical forces on the face related to diet and food processing can account for significant amounts of temporal variation in craniofacial morphology. These trends help explain why central tendencies can shift across time and space without any genetic change or replacement occurring. However, it is wise to remember that there will always be considerable within-population variation around these central tendencies as well.

Perhaps nowhere do we see this problem laid out any better than in Stephen Ousley and R. Jantz's own research on secular trends in populations studied for forensic purposes (1998). They state that due to secular trends in skeletal dimensions from the nineteenth century to the twentieth, comparisons may be "inappropriate" and modern forensic databases should be used (1998:454). These problems result not only from migrations but also "nutrition, medicine, and hygiene, resulting in different patterns of growth, development, morbidity and mortality" (p.
454). Ousley and Jantz also point out that Archaic samples like Indian Knoll cannot be used for ascertaining differences between whites and Indians because of secular changes in the skull height of whites: "the relationships of American Indians and whites has changed: whites now have higher mean skull heights than American Indians" (1998:447-448). If this transition took place in 100 or even a few hundred years, imagine what is capable of taking place in 8,000 years, even if the earliest populations continued as hunter-gatherers for 6,000 years.

Statistical Inference

Owsley and Jantz, as noted, use Mahalanobis's \([D^2]\) statistic to infer relatedness between the early Holocene-Archaic specimens and the late prehistoric/historical samples with which they are compared, and their method for assigning affinity is by using the "typicality probability" as defined in Wilson's oft-cited article (1981); however, our reading and interpretation of Wilson differs from Owsley and Jantz's (2001) in the following ways:

(1) We note that Wilson warns (1981:208) that it is critical to examine normality in each individual measurement, also for multivariate normality (e.g., Mardia 1970; Small 1980), and then to make any necessary transformations, before using the model. This probably should be done given the fact that all but the Great Basin sample contain individuals plotting beyond Gordon Creek Woman.

(2) The probability Wilson infers from the test is 1-p (the probability of obtaining an observation from that population which is more typical than the individual itself). This is a very different emphasis on interpretation than reporting the inverse, p as a probability that this individual belongs to the population. In fact, what the critical level p is, by definition, is the probability that other members of the population distribution will be even more distant. Wilson is careful not to use the probabilities in the way that Owsley and Jantz do because she recognizes that the critical level alone cannot establish that an individual does not belong to a population unless there is only a very trivial chance of more distant members. Conventional significance levels do not apply to an exclusionary use of the probability for classification. To say, for example, there is only a 10 percent, i.e. \(p = .1\), probability that other members of the population would be of more distant morphology is hardly a compelling argument that a given specimen equally distant could not have come from that population. In fact, it says it is quite within the realm of chance (1 in 10) that the individual could be from the population. Only in the case of the Plains does the probability for finding members more distant than Gordon Creek approach the trivial level, which would give confidence to a use of the probability for exclusionary classification. However, for the Great Basin (based on only 26 reference cases) there is still roughly a 4 percent chance, that is 1 in 25, that any individual found that does belong to the population will, in fact, be more distant in morphology than Gordon Creek woman. That is hardly convincing evidence for excluding the remains from even this rather small population sample grouping. A 1-in-25 chance may pose unacceptable limits for Owsley and Jantz, but a quite acceptable chance for an American Indian tribe considering repatriation. (6)

(3) It is possible we have missed something, but nowhere have we been able to find an analysis of the reference population samples in Owsley and Jantz's work. A logical procedure before undertaking these kinds of analyses would be to evaluate the variation among the samples...
themselves. To correctly test Owsley and Jantz's hypothesis we would suggest that they run and evaluate a multivariate classification model (e.g., multinomial logistic regression) that could assess whether between-group variation was sufficient for reliable prediction of group membership and report the possibility of misclassification for each group. Another interesting analysis would be to explore how the addition of one or two specimens to the samples changes their distributions. This would seem to be especially important given Howells's comments regarding how similar American Indian and European groups appear to be in his samples (e.g., 1973:155), and Relethford's (1994) assessment of within- and among-group variation in the Howells data. (7)

Thus, even with these samples selected by Owsley and Jantz, and disregarding the many problems raised with the measurement of Gordon Creek Woman, she can not statistically be rejected from any of the population samples. Moreover, the sample selection and statistical interpretation of the results appears flawed, and Owsley and Jantz's conclusions go well beyond the restrictive limits of the method, not just for Gordon Creek, but for other comparisons as well.

In Jantz and Owsley's recent publication on early North American crania (Jantz and Owsley 2001), they suggest that many early crania do not fit into recent samples. Yet, Wet Gravel Male fits well into as many as five American Indian samples (p. 151), whereas Wet Gravel Female's first five probabilities are completely out of the western hemisphere altogether (p. 151). Does this mean that Wet Gravel Male is an immediate candidate for repatriation, whereas Wet Gravel Female is not? We would suspect that if the methodology was robust and the data adequately informative, Wet Gravel Male and Female would plot more closely to each other. One of Howells's criteria for validating his samples is similarity by sex.

These comparisons yield results and interpretations that are not straightforward and which do not translate well to general audiences interested in these fascinating questions about settlement of the western hemisphere. Van Vark and Schaafsma (1992:235) suggest checking results of Mahalanobis's $[D^2]$ runs for consistency by doing several runs with different sets or subsets of data. In their analysis of Spirit Cave Man from Nevada, Owsley and Jantz (1997) may approximate this procedure when they use subsets of the original data set to look at different components of the cranium. They describe Spirit Cave Man as having the cranial vault of an Ainu, a face breath and vault of an Atayal, the facial forwardness and prognathism of a Norse, and the height, breadth, and facial projection of a Berg (1997). In the final, combined analysis (p. 80), Spirit Cave Man's five closest plots are two European populations (Norse and Zalavar), one Asian population (Ainu), and two Indian populations (Blackfeet and Numic). Again, this would seem to call for the results of a classification analysis. Owsley and Jantz have experimented with the kinds of sophisticated models that we called for in our article (p. 574), but they have not always deployed them in the more complex and "nuanced" ways for which we also called.

Owsley and Jantz charge us with bringing up the false issue of race. Our intent was to focus, as noted, on the initial pronouncements on Kennewick Man by others, and that was where our comments were directed, not at their approach. This is a very complex issue that is further complicated by the professional use of certain terms and the media's imprecise interpretation of those terms, such as the difference between Caucasoid and Caucasian. While there may be some
differences in approaches to classifying biological variation between Owsley and Jantz and us, these are not at issue here.

What Is a Native American?

Owsley and Jantz specifically query how we define a Native American (2001:568). While we use "American Indian" and "Native American" interchangeably in our Forum article, we do regard "Native American" to be a somewhat broader term. "Native American," to us, refers to individuals who trace their ancestry to peoples indigenous to America prior to 1492. It can also include Euro-Americans or others who married into American Indian tribal communities historically, or who were adopted and/or parented children in those communities and are duly recognized as members by those communities. Following Boas and Kroeber, we recognize the independence of "race" (biological affinity), language, and culture, and we regard American Indian "identity" in this context to be primarily a cultural category, not a biological one. While we, as anthropologists, have interests in biological relatedness between populations, we fully understand that our definitions and categories may be different from, and irrelevant to, those of some Native Americans.

There are many theories at present about who composed the indigenous population(s) of America prior to 1492. As noted previously (e.g., Anderson et al. 1997), we find the evidence for a complex process of migrations into the New World compelling. We are also aware of Native American narratives that do not require migration to explain their origins. These will not likely be resolved into any consensus, but regardless of the differences in various theories and narratives, we regard an 8,000-year-old resident of the Americas as a Native American, much as we would tend to regard an 8,000-year-old resident of Britain a Native Briton, even though the genetic origins of the latter might have been in what is now Eastern Europe. So our sense is that, even if the methods for ascertaining differences between populations of modern Homo sapiens could discriminate among those populations, it would probably not be very useful in resolving these debates. And, as with our previous comment regarding Gordon Creek Woman and Kennewick Man, we think Spirit Cave Man and Gordon Creek Woman would not have looked so very different to each other. Moreover, we suspect it quite possible that members of each of their biological communities contributed to the gene pool of contemporary American Indians. This says nothing definitive about tribal affiliation to us, but it might to some.

Our experience in collaborating with Native Americans, and talking with others involved in collaboration, has indicated that there is a strong interest on the part of many in knowing about what the remains can tell all of us about past lifeways, including cultural activities, work, stress, diet, pathology, trauma, and so forth. Obviously, there is much interest also in the effects of colonial contact and subsequent interactions between Euro-americans and Indian-americans. We have not encountered nearly as many interested in having anthropologists tell them who they are, or who they were. Biological anthropologists have a wide array of scientific agendas, and some of us are much more interested in variation insofar as it explains past lifeways and experiences than we are in biological affinity. Like Owsley and Jantz, and as noted previously (Swedlund and Anderson 1999:874), we embrace the opportunity for all parties to have ongoing access to these earliest human remains.
Claims for repatriation will likely be most successful when made on more recent remains, particularly those in which the cultural and biological continuities are fairly unambiguous. It is our opinion that biological tests of affiliation involving craniometrics of a single individual’s remains are fraught with ambiguity if they date from, say, the early Holocene to the present. Comparisons of craniometrics between samples of populations in this time frame afford some improved opportunities for statements on proportional biological distances among groups, but not individual identities. (8)

To summarize, our original Forum essay was aimed at reminding members of the bioarchaeological community that these are complicated times in which to make claims of biological affiliation. Bioarchaeologists no longer have the luxury of unfettered, lengthy investigations of human remains, as we had with Gordon Creek Woman. Current study of both recent (e.g., Kennewick) and past discoveries of early human remains will be subject to the Federal guidelines of the Native American Grave Protection and Repatriation Act (NAGPRA) and requests for repatriation from Native American communities. Federal agencies, scientists, and Native American communities can be expected to have differing interpretations. To our previous point that this places all scientific research into a political climate that cannot be avoided, and that premature claims of ethnic or racial identity can impede meaningful negotiations, we would add that the current methodologies utilized for biological identification of individual remains pose several problems in their implementation and interpretation.

Acknowledgments. We would like to thank the three anonymous reviewers for their useful comments. Also, we thank Douglas Anderton, director of the Social and Demographic Research Institute, University of Massachusetts, for consultation on the statistical methodology issues raised in this reply. We thank Richard Meindl, Department of Anthropology, Kent State University, for comments and suggestions on an earlier draft, and Lynette Leidy-Seivert, University of Massachusetts, for translation of the abstract.

Received January 11, 2002; Accepted June 27, 2002.

Notes

(1.) We could only respond to those published and web-based sources available prior to our submission. One very fair criticism raised by Owsley and Jantz is that referring to the DNA sample sent to the University of California at Davis and our statement that this was done "apparently in hopes of finding haplotypes.... "; we should have said something like "in hopes of finding haplotypes that might suggest the presence of European or Amerindian ancestry."

(2.) It should be noted here that, according to Powell and Rose (1999) and Chatters (2000), the face of Kennewick Man was also in several fragments and required reconstruction, and there are differences of opinion as to the correct measurements.

(3.) We note that in their recent analysis of early North American crania Jantz and Owsley do not include the Great Basin sample for comparison (Jantz and Owsley 2001).
(4.) Because human populations add the complexity of overlapping generations with intergenerational gene flow and exchange, it is even considerably more complex than this implies. Also see Powell and Neves (1999) on this point.

(5.) The example of Harding (1990) given by Owsley and Jantz illustrates that recent population history may be reflected in both blood polymorphisms and morphological variation, but the correlations say nothing about the actual genetic determinants of the morphological variation, or that correlations in geographic distribution result from the same underlying genetic causation.

(6.) We would also note here that Wilson recommends using \([d^2]\) corrected for small sample size rather than \([D^2]\) in the test. This may not make a huge difference in this particular analysis, but perhaps it should be done.

(7.) For example, Howells (1973:55) states that: "Europeans and American indians have a similarity couched in a lack of cranial differentiation and are broadly related in a northern hemisphere complex which also includes Greenland Eskimos and Hawaiians."

(8.) Forensics cases may be an exception as they provide some special requirements and evidence that, when met, permit individual classification, and which, in the right circumstances, may outweigh the chances for misclassification.

References Cited

Anderson, D., A. C. Swedlund, and D. A. Breternitz
Boas, F.
Breternitz, D. A., A. C. Swedlund, and D. Anderson
Chatters, J. C.
Harding, R. M.
Howells, W. W.
Jantz, R. L., and D. W. Owsley
Larsen, C. S.
Mardia, K. V.
Ousley, S. D., and R. L. Jantz
Owsley, D. W., and R. L. Jantz
Powell, J. F., and W. A. Neves
Powell, J. F., and J. C. Rose
Relethford, J. H.
Small, N. H. J.
Steele, D. G., and J. F. Powell
Swedlund, A. C., and D. Anderson
van Vark, G. N., and W. Schaaftsma
Wilson, S. R.