THE EVOLUTION OF EARLY *HOMO*: A REPLY TO SCOTT

A. P. Van Arsdale[1,2] and M. H. Wolpoff[3]

[1]*Department of Anthropology, Wellesley College, 106 Central Street, Wellesley, Massachusetts, 02481*

[2]*E-mail: avanarsd@wellesley.edu*

[3]*Department of Anthropology, University of Michigan, Ann Arbor, Michigan, 48109*

Scott presents a welcome reply to our article, "A single lineage in early Pleistocene *Homo*" (Van Arsdale and Wolpoff 2012). However, Scott's reply mischaracterizes and fails to directly address the hypothesis of a single lineage that we test. Additionally, the approach taken by Scott fails to replicate the methods used in our analysis. As Scott himself suggests, our null hypothesis of a single evolving lineage in early *Homo* remains without refutation. Although many evolutionary scenarios might explain the complex pattern of variation present in the early *Homo* fossil record, the most parsimonious remains that of a single lineage displaying evolutionary change over time.

KEY WORDS: African evolution, fossils, *Homo erectus*, *Homo habilis*, phylogenetics.

Recently, we published an analysis of variation in the early *Homo* fossilrecord from East Africa andGeorgia, suggesting the hypothesis of a single, evolving lineage of Homo cannot be rejected by a parsimonious treatment of the data (Van Arsdale and Wolpoff 2012). Although the fossil record from this time period is relatively rich, it is also complex, and has defied a clear consensus despite the discovery of an increasing number of well-preserved specimens. Despite the lack of consensus, there has been a discernable shift toward models involving multiple, concurrent species of *Homo* for this time period. A recent paper by Antón (2012), in which she identifies multiple morphs of early *Homo* in East Africa and South Africa, is one example of this trend. Our research and its conclusions clearly defy this trend.

Our publication, "A single lineage in early Pleistocene *Homo*: Size variation continuity in early Pleistocene *Homo* crania from East African and Georgia," included within its supplemental materials the largest available set of cranial measurement data for early *Homo* (Van Arsdale and Wolpoff 2012). Our hope in publishing these data was to generate responses to our argument, incorporating the data we provided, as a means of further understanding the critical but complex early evolution of our genus. As such, the reply from Scott (included in this volume) is a welcome development.

Our approach was refutory; we proposed the simplest hypothesis that could explain the craniometric data, including its dating, and we provided novel methods to test it. Our methodology was inclusive in that we devised a way to use the entire dataset, even though each specimen preserved different measurement data. We were able to statistically demonstrate that the simplest explanatory hypothesis of a single lineage changing over time could not be refuted. Scott, in his reply, did two things differently. First, he did not attempt to refute a null hypothesis, but rather asked how many different hypotheses could fit the data. Second, he did not address his question with the data we provided, but rather greatly limited the data used in his analysis. In all, he showed that by severely limiting the data and discarding most of the already albeit limited information available, and by asking a different question, he could demonstrate that several phylogenetic hypotheses could "fit" the remaining data.

Different Questions

The reply from Scott fails to directly address either the question or the key findings we present in our work. We were explicit in stating that our question of interest is whether we can reject the idea of a single, evolving lineage in the early *Homo* cranial

sample considered in our analysis. We began with this hypothesis as astarting point both because it is the simplest, and because of the lack of consensus that currently exists surrounding the phylogenetic pattern of evolution in early *Homo*, particularly when such seemingly enigmatic specimens such as KNM-ER 1805 are included (Aiello and Antón 2012). Our preference was to make as few assumptions as possible in developing the simplest hypothesis, so rather than assume and directly test a specific taxonomic scenario involving multiple lineages in this group, picked from one of many scenarios that have been suggested, we started from a less assumption-laden beginning, including both as broad a set of fossils and as broad a set of measurements as possible.

Scott's paper asks a different question, testing whether a multiple lineage hypothesis could be consistent with a subset of the data we present. Scott's attempt to reject a multiple lineage hypothesis, while related to the question we ask, is semantically and epistemologically distinct from our approach. In Scott's words, "These results *do not reject* the single-lineage hypothesis, but they do indicate that rejection of multiple lineages in the early Pleistocene *Homo* fossil record is premature" (emphasis added). In our article, we never reject the idea that multiple lineages are potentially present in this sample, but rather suggest that the data do not, at present, reject the more parsimonious explanation of a single lineage evolving over time.

Different Assumptions

Although Scott suggests that (presumably our) assumption of *H. habilis* and *H. erectus* coexisting throughout the time period we analyze is not consistent with the current fossil record, multiple sources, including one cited by the author (e.g., Leakey et al. 2012), support such a view. In Leakey et al. (2012), it is argued that *H. habilis* persists through the Early Pleistocene time span of *H. erectus*. Moreover, the taxonomic model for these specimens presented by Scott is just one potential specific scenario for this time period, of which a countless multitude might be put forward and tested.

Scott notes the absence of KNM-ER 42703 from our sample, a potential late-existing *H. habilis* specimen, and therefore one that arguably extends the time period of overlap between *habilis* and *erectus*. The reason it is absent is that the only available published data from this specimen are 12 dental measurements, which are not part of our analysis, and three estimated palate dimensions. These data are not sufficient to meet the minimum measurement requirement of the multivariate approach outlined in our article, thus leading to its exclusion (note that we do include the KNM-ER 42700 specimen published in the same article with a larger set of measurement data). However, the specimen is key evidence for our (above) assertion that specimens attributed to *H. habilis* were found during the time span that specimens attributed

to *H. erectus* were found. We do not ask whether *H. habilis* evolved into *H. erectus*, we ask whether the crania from the time span considered is a single evolving lineage. Accepting for the sake of argument that we took the single evolving lineage as the best unrefuted explanation, it surely would be invalid to pick specimens out of it and name them to different taxa because of their anatomy—that would contradict the assumption. Of course, if we had reason to know that there were two different species in our cranial sample, and were able to unambiguously identify which specimens belonged in each, there would be no point in having this discussion to begin with!

Reduction of Data

We are constrained, as are all paleoanthropologists, by the nature of our data. Few specimens are represented in the fossil record and these might not be indicative of the populations from which they were drawn. Furthermore, these are incompletely preserved and it is often the case that different parts are preserved in different specimens. One paleoanthropologist likened the situation to putting together a jigsaw puzzle when most of the pieces have been thrown away. The key to understanding our approach is that we are not putting together a jigsaw puzzle—we are testing hypotheses about the puzzle.

Yet, rather than making the best use of what is left, Scott leaves out most of the very limited information we have available when he reduces the multivariate approach we employ to a test involving a single variable. In Scott's words, "the use of a single simulated trait modeled on one of the best-represented measures of neurocranial size in the sample probably does not result in the loss of much information." Our dataset (Appendix S1 of Van Arsdale and Wolpoff 2012) included 2921 individual measurements spanning 37 individual fossils. The notion that the reduction from 2921 pieces of data to 21 involves a nonsignificant loss of information is problematic.

Scott's reduction of data involves a reduction in the kind of information available for testing. As Scott suggests, embedded within our data is a complex covariance network between cranial measurements. Our goal was to be as inclusive as possible, thereby including a broad set of specimens, each preserving a different, but overlapping set of measurement data. The nature of our data makes it mathematically complex, and for the most part, mathematically impossible to deeply interrogate this implicit dimensionality within our data. As such, we make no effort to assume or impose an established covariance–variance matrix for the craniometric data we examine, but rather let this property of our data rest on its own. However, simply because we did not probe this aspect of the data does not mean it is not relevant for the conclusions we draw about the data. By reducing the complex, multivariate properties of our data to a single measurement Scott

not only throws out information, but changes what information is being incorporated into his work.

Scott uses the statistical properties of the biasterionic measurement values contained within our sample to construct simulated datasets. These simulated results demonstrate the possibility of observing patterns consistent with multiple lineages with data equivalent to that used in our analyses. This is not surprising, because the more limited the data are, the larger the number of hypotheses they could support. Indeed, across the measurements included within our dataset it is possible to observe a myriad of individual variable patterns.

Scott observes a 21% increase in size between an identified *H. erectus* group (time interval 5) and an earlier identified *H. habilis* group (time interval 1), with both samples having a coefficient of variation (c.v.) of approximately 8. Across our entire sample, using the taxonomic designations outlined by Scott, this later *erectus* group is, on average, 25% larger than the earliest identified *habilis* group, although the *erectus* group has a higher average c.v. (13.1) than the *habilis* group (10.2). Averaged across the entire sample, the range of variability of time interval subsamples, based on Scott's taxonomic assignments, extends from c.v. values of 7.7 (time interval 2, *H. habilis*) to 15.5 (time interval 3, *H. erectus*). Likewise, the size difference between *H. habilis* and *H. erectus* groups Scott identifies also varies across time intervals. If one compares the *habilis* sample within time interval 2, for example, to the *erectus* sample in time interval 4, the *erectus* group is only 7% larger, on average.

These observations are meant to point out that while there is considerable variation that can be parsed from within our dataset, it is not clear why the variability of one trait, even if it is the most highly preserved trait within the sample, can or should stand in for the variation across the entire sample of variables. Moreover, it is unclear why Scott's taxonomic model, which includes changing patterns of variability over time within the designated species groups he proposes is a better explanation of the pattern of variation than our model.

To further emphasize this point, it is worth looking at the metrics Scott uses to generate simulated datasets for the entire dataset, using Scott's proposed taxonomic model. Table 1 displays the time interval subsample c.v., averaged across all available measurements, for Scott's proposed two species model (i.e., *H. habilis* and *H. erectus*) and a single lineage model. Scott's segmenting of specimens between these two taxa serves to create greater variability in the magnitude of intraspecific variation across time intervals. More significantly, while Scott uses the size increase between time intervals 1 and 5 as the basis for his simulated datasets, the overall pattern of size differences using Scott's taxonomic assignments shows a pattern of reducing size differences as the temporal divergence between samples narrows (Table 2). In other words, exactly the pattern one would expect

Table 1. Proposed intraspecific variation across time interval subsamples.

Time interval	*H. habilis* c.v.	*H. erectus* c.v.	Single lineage
1.9–1.8	10.22		10.22
1.8–1.7	7.75		7.75
1.7–1.6	10.86	15.49	12.87
1.6–1.5		9.55	9.55
<1.5		13.08	13.08

Coefficient of variation (c.v.) for each of the time interval subsamples, averaged across all of the available measurement data, based on Scott's taxonomic scheme (columns 1 and 2) and the single lineage model we examined (column 3).

Table 2. Average size difference across time interval subsamples.

	H. erectus, time 3	*H. erectus*, time 4	*H. erectus*, time 5
H. habilis, time 1	1.16	1.15	1.25
H. habilis, time 2	1.08	1.07	1.17
H. habilis, time 3	1.15	1.12	1.17

Relative size increase between the time interval subsamples designated by Scott as *H. erectus* and those assigned to *H. habilis*. Note the overall pattern of reduced relative size increase as the time intervals between the two proposed taxonomic groupings converge.

for a single lineage showing increasing size across evolutionary time.

The strength in our approach is that it made as robust as possible by including measurement data across broad and diverse areas of the skull, as contrasted with an observational account of what happens to one trait. Although our analysis is focused on size variation, the fact that it is multivariate means that aspects of shape variation, while not unpacked in our analysis, are inevitably part of the pattern we observe.

Testing Our Test

Finally, Scott points to the failure of one of our analyses to distinguish the early *Homo* and *Australopithecus boisei* lineages within our sample. This is a misunderstanding of how multiple tests of the same question are combined. It is equivalent to "disproving" that men are larger than women in a populational sample with the observation that the largest woman is bigger than the smallest man. This one result reflects the conservative nature of our approach and the fragmentary reality of the sample available to us, but does not supplant the broader conclusions of our complete set of analyses.

To the broader issue, the limited nature of the fossil record available for studies of early *Homo* make testing power a concern in nearly any analytical approach to this topic. The incorporation

of a broad fossil sample in our analysis is related to this point, but given the different preservation status of our sample, makes a traditional power analysis impossible. To demonstrate the power of our analysis to reject a null hypothesis we included a comparison between early *Homo and A. boisei*. Our analyses reliably distinguish the pattern of variability across time in a sample of early *Homo* mixed with the limited available *A. boisei* fossil record. Simultaneously, our tests fail reject a single lineage model for the sample of early *Homo* on its own.

Conclusions

These distinctions make Scott's work fundamentally different from our own, and thereby are neither a valid response to or nor a rejection of our research. Scott's extreme reductionist approach leaves out most of the available information, does not address our hypothesis in the manner we test it, and fails to suggest a valid alternative approach.

The challenge in viewing Scott's analysis as a reply to our own is that it begins from a different starting point that makes the comparison of our hypothesis tests challenging. Scott presents a specific model of two lineages and then finds evidence for that scenario within our data. Scott's proposed scenario could be supplanted by a multitude of other such scenarios, each of which might also be supported by some fraction of the data we present. But no such approaches would get at the question we address, which is whether a simpler model involving a single, evolving lineage can be rejected given as much information as the fossil record can reasonably provide at this time. We included within our analysis the comparisons with *A. boisei* to show that a rejection of our null hypothesis is possible, even if such a rejection has not been demonstrated here by Scott.

We appreciate the need to understand the opinions and viewpoints that the discoverers of new fossil specimens have about the specimens they discover. It is important to take these into account, as new data become part of our knowledge of human evolution. We hope that the work of Scott and our own article will demonstrate the need to also make systematic observations and measurements of new fossils available, so to continue to generate research and discussion on this and other critical time periods in human evolution.

The current divergence of views on the taxonomic status of fossils associated with early *Homo* reflects a changing tolerance toward specific kinds of analytical assumptions and the associated focus on specific key fossils. Our approach privileges the broad scope of the fossil record and assumptions that recognize our relative weakness in controlling the statistical properties of such samples. Many of our colleagues base their interpretation of the fossil record on assumptions that privilege the kind of statistical control afforded by a limited fraction of the record. Neither approach is inherently correct or in conflict, but we argue that an appreciation for the properties of fossil-generated knowledge favors the more assumption free approach we employ.

LITERATURE CITED

Aiello, L. C., and S. C. Antón. 2012. Human biology and the origins of *Homo*. Curr. Anthropol. 53:S269–S277.

Leakey, M. G., F. Spoor, M. C. Dean, C. S. Feibel, S. C. Antón, C. Kiarie, and L. N. Leakey. 2012. New fossils from Koobi Fora in northern Kenya confirm taxonomic diversity in early *Homo*. Nature 488:201–204.

Van Arsdale, A. P., and M. H. Wolpoff. 2012. A single lineage in early *Homo*: size variation in early Pleistocene *Homo* from East Africa and Georgia. Evolution 67:841–850.

Associate Editor: D. Carrier