Received: 28 May 2022 | Revised: 28 July 2022 | Accepted: 3 August 2022

DOI: 10.1002/ajpa.24608

MEDIA REVIEW

AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY

WILEY

A moving target: Legacy review of The Mismeasure of Man

By Stephen Jay Gould, 1981.

New York, NY: W.W. Norton & Company. 352 pp. ISBN: 0-393-01489-4 (paperback)

1 | INTRODUCTION: A ROADMAP

This legacy review of *The Mismeasure of Man* reflects my thinking about issues early in my career, at the confluence of several pathways I was following as a young scholar that shaped my subsequent thinking. This intellectual history, the way my perspective on how race and human evolution were intricately related, informs my understanding of Stephen Jay Gould's (1941–2002) writing, as I describe below.[1] It was a linear journey for me, reflected in the linear structure of this review, and its conclusion that an incorrect 19th Century conjecture about human races provided support for socially unacceptable conclusions that lasted far longer than the conjecture itself persisted. The review has a beginning from the time before I read the book, a middle as my understanding of it jelled, and an end well after it was published.

2 | BEGINNING: MY BAGGAGE

I began graduate school in anthropology at the University of Illinois, Urbana, in the fall of 1964. My advisor-to-be, Eugene Giles, had been hired to begin his position there at the same time. From my undergraduate background in physics, I fortuitously brought several skills that proved to be very helpful in my approaches to biological anthropology: first of all, a knowledge of physics that included good analytical skills in structural mechanics and associated mathematics, as well as statistics, a practical background in computer programming, and a particularly helpful seminar in the philosophy of science that in retrospect was one of the two most influential contributions to my undergraduate education. This was taught in the Department of Philosophy by Joseph Agassi, a student of Karl Popper. The other, also a seminar, was on the fossil evidence for human evolution, taught in the Anthropology Department by Donald Lathrap.[2] Basically, Agassi addressed how to do science and Lathrap focused me on what science I wanted to do. In retrospect, I am quite sure that Lathrap's seminar opened the door (intellectually, if not practically) for me to begin graduate work at Illinois in biological anthropology, especially since during the following year, my senior year, he asked me to lead his seminar for 2 weeks while he was away at the AAA meetings. This was an honor I would not have expected as an undergraduate student in Physics.

In graduate school, Giles, by then my new advisor, was insistent on supplementing my anthropology requirements with biology, human anatomy, and population genetics. With my background, I became particularly interested in his ideas about the role of allometry in the evolutionary process; 8 years earlier, he had published a paper on cranial allometry in the great apes (Giles, 1956). His proposal, that gorillas could be thought of as chimpanzees that continued further along the same allometric curve, struck me as a possible explanation for how the gracile and robust australopithecines from South Africa might be related.[3] In the 1960's the problem of relating these australopithecines was both constrained and simplified by the fact that the only sizable samples were from a few caves in South Africa that were quite close to each other.

3 | ALLOMETRY

Then, in (Gould, 1966) there was a substantive analysis "Allometry and size in ontogeny and phylogeny," detailing his thoughts and interpretations of allometry and allometric relationships, that was very helpful in organizing my own thoughts on allometry. Eleven years later this was followed by *Ontogeny and Phylogeny*,[4] that served as a good introduction to *The Mismeasure of Man* for me. It set the stage for what was to come in *The mismeasure of man* by addressing the importance of brain size in human evolution, and at least implicitly recognizing the contradictions between how we treat brain size differences across evolutionary time and how we treat brain size differences freezing time and looking across biological space. Allometry was a persistent theme in Gould's earlier writings, one key reason why they caught my attention and challenged my thinking.

Gould was a biological scientist, historian, and philosopher of science, as well as a skilled popularizer of evolution (Shermer, 2002), first and foremost because he was a very good writer. In the 1977 book, he defined allometry as (p. 479): "change in shape correlated with increase or decrease in size." This was deeply unsatisfying because allometry had always been, and up to this point in time remained, an observation of a certain kind of bivariate relationship without a really adequate theory to account for it.[5] I expected that there was an important role for allometry in early hominid[6] evolution but lacked the evidence for it and the theory to explain it.

I took my first course in computer programming in 1962 as a sophomore in physics at the University of Illinois' Digital Computing Laboratory,[7] famously known as the birthplace of HAL 9000. Later, as

a graduate student, I approached understanding how allometry worked by looking broadly at patterns of allometry in human crania, taking advantage of the fact that allometry is expressed when the log–log plots of independent and dependent variables form a straight line, or close to it. I used measurements of human crania published in *Biometrika* during the first half of the 20th century, in a "brute force" examination (as we called it then), printing out log–log plots of each cranial measurement compared with the logs of every other one. The point of this exhaustive approach was to look through the printouts for patterns. Did the same combinations show allometry in all or most of the human samples? Could common patterns of human allometry show how chimpanzee and gorilla cranial dimensions are related by allometry (Giles' hypothesis)? Would that pattern explain australopithecine relationships? But the data were having none of it. The brute force approach showed me only that that there was no overarching pattern of allometry for the modern human cranial samples I examined. Bivariate plots of logs that *were* linear were rarely the same across the samples, from one comparison to another. Exhaustive brute force examination revealed nothing.

By the time *The Mismeasure of Man* was published, the australopithecines had come to include numerous, better preserved, taxonomically diverse, East African remains. The question of possible relationships based on allometry became much more complex because for the first time there were associated crania and postcranial remains. I ultimately did publish a paper on allometry Wolpoff (1985), but it had nothing to do with early hominid relationships. In hindsight, the most important thing I learned was never to approach data without the framework of a theory they could test. The second most important thing was an interest in Gould's work.

4 | RELATIVE BRAIN SIZE AND RACE IN *ONTOGENY AND PHYLOGENY*

In his 1977 book, an elaboration of the paper he published earlier, Gould made the case that 19th Century racism arose from the evolutionary science of the period, wherein one could learn that "children of higher races (invariably one's own) are passing through and beyond the permanent conditions of adults in lower races. … The 'primitive-as-child' argument stood second to none in the arsenal of racist arguments supplied by science to justify slavery" (p. 126; Gould, 1977). It should be noted that *this argument inexorably*[8]*links allometric curves and growth curves*.

Gould's discussion was about allometry and brain size evolution, not so much about brain size variation. And brain size evolution and variation were quite difficult to reconcile because *it was unclear whether allometric curves of brain–body size actually were the same as growth curves*.[9] Brain size variation was important in many discussions of races and their "evolutionary status." In fact, in Gould's time probably nothing was as heatedly discussed as the meaning of differences in brain sizes between human populations, because these were suspected to reflect differences in intelligence[10] (see below).

5 | MIDDLE: *THE MISMEASURE OF MAN*

The Mismeasure of Man (Gould, 1981)[11] addressed a fundamental question about human groups; mainly, how scientists had attempted to rank them. The object actually being ranked, one way or another, was the human brain, and the groups in question (at least historically) were usually human races. The basis of ranking could be wildly variable but most often it was about human intelligence. This was particularly important to Europeans in the second half of the second millennium, their most active period of colonialism.

The deeply seated hypothesis in *The Mismeasure of Man* is that across human groups (usually races) brain size, intelligence, and status of civilization, formed a kind of *triangle of relationships*. Gould's (unstated) attempt was to critically examine the strength of each of its sides. In doing so, Gould focused on the meaning of differences recognized in intelligence (as measured by IQ) and brain size (as measured by the cranial capacity of museum specimens) among human groups, sometimes races and sometimes not. In a nutshell, this came to discussing Arthur Jensen (1923–2012), who ranked races by their IQ, and Samuel Morton (1799–1851), who ranked races by brain size.

Neither was valid, and what else they had in common was more than ranking races, it was (1) their assumption that what they thought of as races *could be* ranked, and (2) their use of inappropriate measurements to do so. Allometry was dropped from the discussion. Gould no longer mentioned allometry in *The Mismeasure of Man*, save for a single footnote, and the status of civilizations proved to be too amorphous to deal with (in any event, how civilizations were ranked was obvious to many Europeans, especially at the height of their colonization activities).

The book was written at a time when the relation of IQ and race had become a red-hot topic involving academia, science, and the public space. Brain size differences between populations were uncritically taken to imply differences in intelligence, although when Gould wrote *The Mismeasure of Man*, the supposition had rarely been examined statistically (excepting Van Valen, 1974), and then, with inconclusive results. Focus on the IQ was at least in part because of an influential but controversial 1969 paper by Jensen in the prestigious *Harvard Educational Review*, and supporting it, ongoing publications by several other authors. These claimed human intelligence was so strongly inherited that improving the intellectual environment during schooling could not compensate for the (assumed) larger innate genetic differences between human races.

How did that work out? Kenneth Beals (1987, p. 59) contended: "normal variation in human brain size has no more significance for mental ability than do randomly selected anthropometric traits." But Jensen and Johnson (1994) went on to examine the relationship between race and sex differences in head size and IQ (their wording) and concluded (p. 309): "Within each race and sex group, IQ is significantly correlated with head size, [with] age and body size having been partialed out." The situation was muddled before and after Gould wrote on it, and I (with many others) would venture to suggest that the problem was with IQ; what it means, or does not mean.

Even without these and other follow-up studies, Jensen's thinking ran contrary to the direction taken by the U.S. Congress in its creation of *Head Start* in 1965, and its outgrowth, the *Children's Television Workshop*. But this direction was not without its critics, and Jensen's research was mentioned in critiques.[12] *Children's Television Workshop* began broadcasting in 1969. Renamed *Sesame Street*, the children's program was based on a very different assumption than Jensen proposed about the relation of nature and nurture. Various aspects of this argument are discussed in *The Mismeasure of Man*, which was quite critical of Jensen's writings. It is telling that Jensen did not like Gould's book. This was later reported by Gould (1996, p. 45), who I believe seemed pleased.

Gould (1981) provided quite a bit of discussion critical of Jensen's research, stoking the fires of controversy. But his aim was well beyond this controversy. Writing *The Mismeasure of Man* created the opportunity to detail his views on human brain size: its evolution, its variation, and its relation to race and intelligence, in the context of the history of science. In other words, his forté. To clearly explain his views on this relationship, Gould approached brain size variation in human groups indirectly, in a perspective that avoided repeating his aggressive position on Jensen. I think one of the main reasons *The Mismeasure of Man* is still cited by many anthropologists today comes from this discussion of brain size and race, and not because of his critique of Jensen.[13]

The vehicle he used for this discussion was Morton's cranial studies.[14] Curiously, despite the heated discussions about race during these and later decades, Gould's considerations of Morton's research in *The Mismeasure of Man* were not mentioned in initial book reviews published in *Science* (Samelson, 1982), *Nature* (Blinkhorn, 1982), or even in *Human Biology* (Durham, 1984). What was it in Gould's writings that was passed over at first, but later came to provoke so much debate?

The first reviews of *The Mismeasure of Man* might have overlooked its discussions of Morton and his research because only *3 years earlier* Gould had published a more specific account of Morton's ranking of human races by cranial capacity in the journal *Science* (Gould, 1978). I had read these out of chronological order. It was later that I realized the earlier *Science* publication would have been more rigorously reviewed than the book, with its more detailed understanding of history. My discussion below relies on both publications.

Morton was an early polygenist, publishing on his cranial collection in *Crania Americana* (1839, 1849). Polygenism is an explanation for the then unquestioned assumption that human groups are biologically different, stemming from the contention that human races are different "types," or even species.[15] In polygenism's pre-Darwinian form "early polygenists [like Morton] viewed races as independent biological units, like species, created separately by God" (Wolpoff & Caspari, 2013, p. 322).

Morton ranked races by the sizes of their brains. He "had a hypothesis to test: that a ranking of races could be established objectively by physical characteristics of the brain, particularly by its size" (Gould, 1981, p. 51). This meant cranial capacity.[16] His choice of cranial capacity as a key metric for comparing races seemed to be a reasonable one for the time. The brain was understood to be the seat of humanity, and other more indirect measurements of it only reflected part of the cranial vault, were less straight-forward to define, and were more difficult to replicate.[17]

From Morton's understanding of history, the short time available was not sufficient to account for the differences he observed among races unless those differences were present from the beginning. Thusly, Morton (1839, p. 88) proposed that "human races must have been separate from the start," referring to the time of creation. This implied that races had different origins, independent creations.[18] Seemingly supported by the science of the time, these conclusions had great political implications in the 19th century as they directly informed the debate about human enslavement.

But today we would describe this as "bad science"; not necessarily bad because it was evil (although it was in many cases), but bad because it was both wrong and led to socially unacceptable assertions. Refutation is the mechanism of correction or rejection for incorrect hypotheses in science, and of course, scientific conjectures can be incorrect whether or not the science can be (or has been) used to promote the socially unacceptable. But the fact is that refutations can be difficult to establish, and even more difficult to defend (Kuhn, 1962), so that bad science could persist for a very long time. This review of *The Mismeasure of Man* exemplifies how an incorrect hypothesis about human races supported socially unacceptable conclusions even longer than the hypothesis itself persisted.

Gould (1981) was justifiably critical of Morton's discussion of race, especially over how races could be validly compared and what the comparisons implied. He found serious issues in the ways race was defined,[19] in how racial variation was addressed by cranial capacities, and especially in the inaccuracies he found in Morton's supporting tables. Gould addressed each of these in detail. And he did more; he examined the supporting data (1978), as described below.

Before there were scanners and the computers that could interpret the scans, cranial capacity was determined by a physical process. A cranium, with soft tissue removed (the usual condition of crania in museums), was held or propped upside-down, bird seed or its equivalent that was small and spherical[20] was poured into the cranial cavity through the foramen magnum. Since the inside of a cranium is an irregular surface, gently shaking the cranium from time to time assured it was filled. The seed was then poured into a graduated cylinder and the total internal volume recorded. This was repeated several times and the average was taken as the cranial volume or capacity. I have been teaching osteology since 1971, and until the 21st century this is how I explained the measurement of cranial capacity.

In Morton's ranking of races, based on their cranial capacities, there are several places where Gould proposes that objectivity was or could have been lost.[21] Here, I discuss two key ones. The first is process. Were all the crania filled completely? Were the seeds equally loose in the crania or were they periodically compressed in some, perhaps unconsciously, as the cranium was being filled? The idea is not that this might have been done purposefully, but rather that it reflected an unconscious attempt to meet expectations. For instance, in a population expected to have larger cranial capacities, perhaps

there was more shaking, and more unconscious compression, as the crania were filled.

The second place objectivity could have been lost comes from the procedure he used in calculating the mean capacity for samples, in cases when the samples were made up of subsamples of different sizes. Neandertals[22] provide a good example of this problem. When I began my first anthropology position, in 1968, it was widely understood in paleoanthropology that the Neandertal sample had bigger brains than modern humans, and this still often is said to be the case. It is one of the few "facts" from paleoanthropology that has permeated into public science (and science fiction). And it is correct if Neandertal cranial capacity is calculated by simply taking the average of all known Neandertal crania. This can be expected to be quite accurate because there are many Neandertal crania complete enough to allow it. But *accurate* is not necessarily *correct*.

It turns out that postcranial remains demonstrate the existing Neandertal sample does not have an equal number of males and females, the number of males is much larger. This might be a chance occurrence that reflects the small size of the sample, but in a resampling study Lesnik and Sams (2014, p. 463) "fail to reject the null hypothesis that Neandertal burials do not deviate significantly from the expectation of random draws from a population with a balanced sex ratio." Yet, even if it cannot be shown that the underlying sample deviates significantly from an even sex ratio, the fact that *the actual sample has an unbalanced sex ratio* will alter calculations of any sex-related feature, such as cranial capacity. Neandertal males have, on average, a larger mean cranial capacity than females (this is usual in human populations). When Neandertal cranial capacity is not calculated as the average of all Neandertals, but rather as the average of the male mean and the female mean, Neandertals do not demonstrate a larger cranial capacity than modern humans (and see VanSickle et al., 2020).

Morton's human sample is not made up of equal sized subsamples,[23] and it *is* a problem. The problem comes from the possibility that bias might have influenced the picking and choosing of which procedure was used (Gould 1981, p. 60), because this was an opportunity for the expected result to influence how that result was determined.

The presence of unequal subsample sizes means there are choices to be made between different approaches when calculating the average for the whole sample, and we may ask why Morton made his. Gould did ask this, and his answer for the human populations Morton studied is bias.[24] Gould proposed four explanations for (what he described as) Morton's "finagling"[25] (pp. 68–69, reproduced here in abbreviated form):

1. Favorable inconsistencies and shifting criteria: Morton often chose to include or delete large subsamples in order to match group averages with prior expectations.
2. Subjectivity directed toward prior prejudice: Morton's measures with seed were sufficiently imprecise to permit a wide range of influence by subjective bias; later measures with shot, on the other hand, were repeatable, and presumably objective.
3. Procedural omissions that seem obvious to us: Morton was convinced that variation in skull size recorded differential, innate mental ability. He never considered alternate hypotheses; [for instance sex or body size]
4. Miscalculations and convenient omissions: All miscalculations and omissions Gould detected are in Morton's favor. Morton rounded the Negroid Egyptian average down to 79, rather than up to 80. He excluded a large Chinese skull and an Eskimo subsample from his final tabulation for Mongoloids, thus depressing their average below the Caucasian value.

Yet, Gould wrote that he did not consider there was a purposeful strategy Morton used to support his biases: "through all this juggling, I detect no sign of fraud or conscious manipulation. Morton made no attempt to cover his tracks and I must presume that he was unaware he had left them (p. 69, italics mine)." Morton's bias, in other words, was *unconscious*, and the implications of Morton's unconscious bias have reverberated to the present.

6 | REVIEWS OF GOULD

In the first of what became a slowly unfolding discussion of Gould's analysis of Morton following the earlier book reviews in major science journals, Michael (1988) compared Morton's, 1839 and 1849 works to each other and concluded: "Morton's tables contain miscalculations and omissions of data, but his 1849 data are reasonably accurate and there is no clear evidence that he doctored these tables for any reason. … I do not argue that ostensibly objective research may not sometimes be affected by the unconscious desires of the researcher; I simply suggest that the work of Morton is not a clear example of this phenomenon." (p. 354). Most of the discussions that followed agreed with Michael's assessment.

In (Vincent Sarich and Frank Miele, 2002) writing just prior to Gould's 2002 death, the authors treated Morton's work favorably, for the most part, and were critical of Gould's interpretations and explanations of it. Many have rejected the arguments Sarich and Miele advance over this and other issues on the grounds that their book supports racist views. Sarich and Miele declared that Gould believed "race *cannot* exist"[26] (p. 8), and that this denial underlay how Gould misrepresented Morton's research. Their claim was that the errors and faulty procedures Gould found in Morton's cranial capacity studies could be identified by remeasuring and reanalyzing Morton's sample. Their conclusion was "the results [of such remeasuring] show that any errors were Gould's, not Morton's" (p. 72).

Brace (2005) was critical of *The Mismeasure of Man*, arguing that Gould's findings of unconscious bias in Morton's research in actuality reflected biases that Gould held. Mann (2009) reviewed these issues and also rejected the interpretation that Morton *unconsciously* manipulated his data to support polygenism.

And then there was the paper by Lewis et al. (2011) that provoked its own set of reactions. This work presented a substantive discussion of the role of bias in science, and specifically in the science we are discussing, centering on the issue of whether Morton had consciously or unconsciously finagled his data to support his polygenic

interpretation of human history. The authors concluded (p. 6): "Morton's methods were sound, and our analysis shows that they prevented Morton's biases from significantly impacting his results." In other words, whether or not his finagling was conscious, the finagling did not significantly alter his results.

The same year there was a *New York Times* review of the Lewis et al. paper that was also critical of Gould (Wade, 2011) and an unsigned *Nature* editorial (Anonymous, 2011) discussing the Lewis et al. paper. This latter commentary was also highly critical of Gould: "at a minimum, Gould's staunch opposition to racism, and desire to make an example of Morton, may have biased his interpretation of Morton's data, opening Gould to charges of hypocrisy" (p. 419).

However, Weisberg and Paul (2016) did not agree that it was Gould's analysis of Morton that was biased. They pointed out: "although Lewis *et al*. found Morton's shot-based measurements to be accurate, Gould had already accepted this. Indeed, Gould had to assume that Morton's shot measurements were accurate, as he relied on them in his own analysis. Gould never made, nor did he ever claim to make, nor did he have any reason to make any measurements himself. Gould's argument depends on the difference between the two sets of measurements. Thus, as a matter of logic, there is no way that the results of Lewis et al.'s remeasurement program could be used to adjudicate the issue of who was biased" (p. 2).

7 | END: BIAS IS A MOVING TARGET

The assertion that Morton unconsciously manipulated his cranial capacity data to reflect a racist bias, whether in the way he collected his data (shot-based) or the way he calculated sample statistics from his data (not correctly accounting for differences in the subsample sizes that made up each sample), *was rejected* by just about everybody who wrote on the topic. But bias is a moving target, and some of these writers asserted that there was actual bias but it was to be found in Gould. What is most surprising to me is the reliance of key arguments, with important implications that extend to the present, *on what is essentially an untestable hypothesis about the mental states of past scientists*.

In my view, it is likely that the controversy Gould raised about Morton was not resolved by the 35+ years of discussions that followed his 1981 book *because it could not be*. Moreover, while Gould's criticisms were of Morton's 19th century work, they raised questions about bias that still deserve unambiguous answers in the 21st. It was clear to Gould and authors who commented on him, and I agree, that Morton was a racist by the understanding of his time and of ours (Lewis et al., 2011; Mann, 2009). This is not at issue, and I do not think it ought to be. Despite this, I also think it is important to understand Morton's work in the context of his own times. In this, I do not think that Gould was fully successful. Gould assumed Morton's racist conclusions were the result of bad science. But the problem was not so much in the science of the time as it was in Morton's racist interpretation of it.

To summarize, in 1981 Gould wrote about Jensen, who ranked races by IQ, and Morton, who ranked races by brain size, both addressing human intelligence. But it turned out that IQ is mostly or totally unrelated to brain size, neither was a useful parameter of human variation, and specifically neither measured human intelligence. What they had in common had nothing to do with ranking races, it was their use of an inappropriate approach to answering a racist question, or in other words, they were both *mismeasurements*.

8 | LET'S TRY IT ONE MORE, ONCE![27]

Extending beyond this legacy review, to better comprehend the entirety of what has taken place I propose it is also important to think about our times. Jensen's arguments about racial differences in intelligence did not survive a half century of *Sesame Street*. To understand Morton's work in the context of our times is more difficult because of the tenuous link, if it is a link at all, between brain size and any measure of intelligence. Furthermore, I suggest that it is not possible to adequately evaluate *The Mismeasure of Man* without considering the vantage point of biological anthropology today (Marks, 2021), which is quite different from the times when Gould's book was written, or of the reviews discussed here. Weighing heavily on this is the recognition that the human skeletal collections themselves, perhaps even more than what has been written about them, have now become problematic in how they reflect racism and its consequences (Mitchell, 2021). This, of course, incudes Morton's collection that was the basis of his research reviewed in *The Mismeasure of Man*.

AUTHOR CONTRIBUTIONS

Milford H. Wolpoff: Writing – original draft (lead).

Milford H. Wolpoff

Department of Anthropology, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Milford H. Wolpoff, Department of Anthropology, University of Michigan, Ann Arbor, MI 8109-1107, USA.
Email: wolpoff@umich.edu

ORCID

Milford H. Wolpoff https://orcid.org/0000-0002-6218-4963

ENDNOTES

[1] Later, and further matured, this led to additional publications (e.g., Wolpoff & Caspari, 1997).

[2] An archeologist with broad interests and experiences, as was characteristic of many Harvard graduates in those days.

[3] In the mid-1960's body size of the two australopithecine samples was poorly established and it was assumed from postcanine tooth size that the robusts were bigger. But the significant significant problem for any kind of allometric analysis was that even 10 years later "there is not a single early hominid cranium with a known cranial capacity that is

associated with any postcranial material useful for determining body size" (Wolpoff & Brace, 1975, p. 62). Later, it became clear that their sizes were similar, and ranges largely overlapped, which meant that allometry could not have played a significant role in their variation.

[4] The expanded text provided an enlarged view of many of the issues in the first edition, and in many cases commented on how the first edition was received.

[5] Until Russell Lande (1979) outlined a predictive model for allometry as populational change over time when an independent and dependent variable are genetically correlated, and selection on the independent variable results in changes in the dependent variable.

[6] Why do I use the term hominid, and not hominin? Fundamentally, it is because like ape or monkey, hominid is not a taxonomic term. Recognition that humans and chimpanzees are sister groups required some taxonomic revisions for humans and the great apes since Hominidae and Pongidae were no longer valid family names. This is because valid taxa must be monophyletic (a taxon and all its descendants), and it was recognized that the last common ancestor of chimpanzees, gorillas, and orangs (all Pongidae) is also an ancestor of humans. The solution was to redefine the taxa. The Hominidae became the family that includes the great apes, the human line, and all the descendants of their LCA. Notice that I can use humans and apes in this discussion because their definitions have not changed. Like the words hominid and pongid they are not taxonomic terms. The problem comes if hominid and pongid are treated as if they *were* taxonomic terms. Then changes in phylogeny would require changes in their meaning. This would be confusing and misleading because two widely used and understood categories, 'hominid' and 'pongid' would be retained, but with very different meanings than they had before. Reading an older literature with these terms would evoke different meanings than their modern use, but older paper publications cannot be modified to have warnings about this problem so there is a real danger of misunderstandings. But while the phylogeny, and thus the taxonomy, *has changed*, terms that are not taxonomic *do not necessarily also need to be changed to conform*. We use "ape" to identify a group of primates that is not a taxon anymore. "Hominid" is like "ape," it can retain its older and widely known meaning just as "ape" has, and there also is no corresponding taxon.

[7] I think I am the first in biological anthropology to include tables and charts in my 1969 dissertation that were direct printouts from a mainframe computer (I had to purchase a full ream of special paper with sprocket holes to allow this, what remained of the ream (most of it) became a source for unnumbered children's drawings over subsequent years). Unfortunately, I could not use computer printout for the dissertation text as we do today, as lower-case letters were not available yet.

[8] But incorrectly.

[9] Yet, one can appreciate why this was an ongoing problem by simply considering the title of a classic publication on allometry, Huxley's (1932) *Problems of Relative Growth*.

[10] Gould (1981) does not cite Van Valen (1974), one of the few earlier papers that provides a statistical discussion of this issue.

[11] I would advise the reader that the introduction to the 1996 second edition be read after the 1981 first addition. It is illuminating!

[12] "Research on results of the Head Start program has been mixed. The scathing 1969 Westinghouse report (Westinghouse Learning Corporation & Ohio University 1969); *the Jensen (1969) article, which stated that 'compensatory education has been tried, and it apparently has failed'* (author's bold; cited in Blythe and Hinitz, 2014, p. 96) …"

[13] Which however correct, is quite dated now.

[14] The cranial collection he put together for these (and other) studies was unprecedented at the time. It became known as "Morton's Skull Collection." Most of it now housed at the University of Pennsylvania Museum of Archaeology and Anthropology.

[15] Morton believed that some of the dog breeds, like some human races, were separate species (Gould, 1981, p. 52).

[16] Cranial capacities (cc) were a logical way to study human variation. Volume is easily defined, it is not subject to different definitions, and it varies within and between human populations.

[17] Nevertheless, Morton did report other cranial measurements and observations

[18] A key tenet of pre-Darwin polygenism.

[19] In some cases, as different as species, and see note 10.

[20] In Morton's time, small uncompressible spheres such as lead shot were also used (beginning a long discussion about whether seeds or shot gave the most consistent results). The fill later became acrylic balls and even BBs. Gould argued that Morton's (1839) seed-based results were inconsistent with his 1849 shot-based data and suggested this was additional evidence of bias (but see Weisberg & Paul, 2016).

[21] This does not mean the process is inevitably subjective, but rather that if there is reason to suspect that objectivity has been lost, these are places to look for it where that could have happened.

[22] To explain my spelling of Neandertal when referring to a site and the human remains found there (questioned in review), I spell "Neandertal," used in this way, without the "h"; that is, not as Neanderthal. This is why. At the time a Neandertal was first named, in 1856, the name of the site where it was found was spelled Neanderthal, and this became the name of the human found there. In 1904, German spelling was made more consistent with normal pronunciation. For one example "thal," the German word for valley, became "tal" because what had become a silent h was dropped. This changed the spelling of the site name, and the human's name. But Neanderthal was also named as a species "Homo neanderthalensis" several years later (but before 1904) and the spelling of a species name cannot be changed because the rules of taxonomic nomenclature do not allow it. This is how we got stuck with "Australopithecus," a mixture of a Greek and Latin root. In any event, my spelling of the term Neandertal reflects my assessment that it refers to a site and the group of people found there and elsewhere, but that *the Neandertal group is not a taxon*. The two spellings of Neandertal are pronounced the same (silent h).

[23] For instance, as Gould reports (p. 57) Morton's Native American sample "belong[s] to many different groups of Indians; these groups differ significantly among themselves in cranial capacity," and each group was different in size.

[24] See Brace (2005). Here, I cite Gould (1981, p. 65): "The more 'inferior' a race by Morton's a priori judgment, the greater the discrepancy between a subjective measurement, easily and unconsciously fudged, and an objective measure unaffected by prior prejudice."

[25] Gould's term for whatever it was that Morton might have done.

[26] Gould's (1996) understanding of race is clearly displayed in his second edition. In my opinion "do not exist" would have been a better description of his thinking than "cannot exist," which was actually used by Sarich and Miele.

[27] Count Basie, on the occasion of asking to have the ending of *April in Paris* repeated a second time, as preserved on the recording.

REFERENCES

Anonymous. (2011). Mismeasure for mismeasure. *Nature*, *474*, 419.

Beals, K. L. (1987). Problems and issues with human brain size, body size and cognition. *Homo*, *37*, 148–160.

Blinkhorn, S. (1982). What skulduggery? *Nature*, *296*, 506.

Blythe, S., & Hinitz, F. (2014). Head start: A bridge from past to future. *Young Children*, *2014*, 94–97.

Brace, C. L. (2005). *"Race" is a four-letter word: The genesis of the concept*. Oxford University Press.

Durham, N. M. (1984). Book review of *the Mismeasure of Man* by Stephen Jay Gould. *Human Biology*, *56*(4), 795.

Giles, E. (1956). Cranial allometry in the great apes. *Human Biology*, *28*(1), 43–58.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews*, *41*, 587–640.

Gould, S. J. (1977). *Ontogeny and phylogeny*. Harvard University Press.

Gould, S. J. (1978). Morton's ranking of races by cranial capacity. *Science*, *200*, 503–509.

Gould, S. J. (1996). *The mismeasure of man* (2nd ed.). Norton.

Huxley, J. S. (1932). *Problems of relative growth*. Dover.

Jensen, A. R. (1969). How much can we boost I.Q. and scholastic achievement? *Harvard Educational Review*, *39*(1), 1–123.

Jensen, A. R., & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence*, *18*(3), 309–333.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain-body size allometry. *Evolution*, *33*, 402–416.

Lesnik, J. J., & Sams, A. J. (2014). Using resampling statistics to test male interment bias: Applications for looted and commingled prehistoric remains in Peru and the reassessment of Neandertal burials. *PaleoAnthropology*, *2014*, 463–469.

Lewis, J. E., DeGusta, D., Meyer, M. R., Monge, J. M., Mann, A. E., & Holloway, R. L. (2011). The mismeasure of science: Stephen jay Gould versus Samuel George Morton on skulls and bias. *PLoS Biology*, *9*(6), e1001071.

Mann, A. (2009). The origins of American physical anthropology in Philadelphia. *Yearbook of Physical Anthropology*, *52*, 155–163.

Marks, J. (2021). On demarcation. *History of Anthropology Review*, *45*. https://histanthro.org/news/observations/on-demarcation/

Michael, J. S. (1988). A new look at Morton's craniological research. *Current Anthropology*, *29*(2), 349–354.

Mitchell, P. W. (2021). Editor's introduction: The Morton cranial collection and legacies of scientific racism in museums. *History of Anthropology Review*, *45* https://histanthro.org/news/observations/editors-introduction-morton/

Morton, S. G. (1839). *Crania Americana; or, a comparative view of the skulls of various aboriginal nations of*. North and South America.

Morton, S. G. (1849). Observations on the size of the brain in the various races and families of man. *Proceedings of the Academy of Natural Sciences, Philadelphia*, *4*, 221–224.

Samelson, S. (1982). Intelligence and some of its testers. *Science*, *215*, 656–657.

Sarich, V., & Miele, F. (2002). *Race. The reality of human differences*. Westview.

Shermer, M. B. (2002). This view of science: Stephen jay Gould as historian of science and scientific historian, popular scientist and scientific popularizer. *Social Studies of Science*, *32*(4), 489–524.

Van Valen, L. M. (1974). Brain size and intelligence in man. *American Journal of Physical Anthropology*, *40*(3), 417–424.

Van Sickle, C., Cofran, Z., & Hunt, D. (2020). Did Neandertals have large brains? Factors affecting endocranial volume comparisons. *American Journal of Physical Anthropology*, *2020*, 768–775. https://doi.org/10.1002/ajpa.24124

Wade, N. (2011). Scientists measure the accuracy of a racist claim. *New York Times*, D4.

Weisberg, M., & Paul, D. B. (2016). Morton, gould, and bias: A comment on "The mismeasure of science". *PLoS Biology*, *14*(4), e1002444.

Wolpoff, M. H. (1985). Tooth size - body size scaling in a human population: Theory and practice of an allometric analysis. In W. L. Jungers (Ed.), *Size and scaling in primate biology* (pp. 273–318). Plenum.

Wolpoff, M. H., & Brace, C. L. (1975). Allometry and early hominids. *Science*, *189*, 61–63.

Wolpoff, M. H., & Caspari, R. (1997). *Race and human evolution*. Simon and Schuster.

Wolpoff, M. H., & Caspari, R. (2013). Paleoanthropology and race. In D. Begun (Ed.), *A companion to paleoanthropology* (pp. 321–338). Wiley-Blackwell.