EVOLUTION & DEVELOPMENT 16:3, 166–178 (2014)
DOI: 10.1111/ede.12077

required pp 171-177

Remeasuring man

Michael Weisberg*
University of Pennsylvania, Philadelphia, PA, USA
*Author for correspondence (e-mail: weisberg@phil.upenn.edu)

SUMMARY Samuel George Morton (1799–1851) was the most highly regarded American scientist of the early and middle 19th century. Thanks largely to Stephen Jay Gould's book *The Mismeasure of Man*, Morton's cranial capacity measurements of different races is now held up as a prime example of and cautionary tale against scientific racism. A team of anthropologists recently reevaluated Morton's work and argued that it was Gould, not Morton, who was biased in his analysis. This article is a reexamination of the Morton and Gould controversy. It argues that most of Gould's arguments against Morton are sound. Although Gould made some errors and overstated his case in a number of places, he provided prima facia evidence, as yet unrefuted, that Morton did indeed mismeasure his skulls in ways that conformed to 19th century racial biases. Gould's critique of Morton ought to remain as an illustration of implicit bias in science.

INTRODUCTION

Samuel George Morton (1799–1851) was the most highly regarded American scientist of the first half of the 19th century and one of the founders of the field of Physical Anthropology. Today he is primarily remembered as a target of attack in Stephen Jay Gould's *The Mismeasure of Man*. Gould argued that Morton's work contained "a patchwork of fudging and finagling in the clear interest of controlling a priori convictions" (Gould 1981, 54). According to Gould, Morton manipulated his measurements of the crania in his collection so that his conclusions would conform to his racist expectations.

Following the publication of *The Mismeasure of Man*, Morton and his work stayed out of public view for another 30 years or so. But in 2011, a team of physical anthropologists associated with the University of Pennsylvania Museum of Archeology and Anthropology (hereafter Lewis et al.) published an extensive reanalysis of Morton's and Gould's work. They reported that "Our results resolve this historical controversy, demonstrating that Morton did not manipulate data to support his preconceptions, contra Gould" (2011, abstract). The publication of this article was followed by many condemnations of Gould for his inaccurate assessment and unfair treatment of Morton. For example, in a *Psychology Today* blog entry, Robert Trivers wrote that Gould was "a blow-hard fraudulently imputing fraud, with righteous indignation, coupled with magnanimous forgiveness for the frailties of self-deception in others" (Trivers 2012).

This article is a re-examination of the Morton and Gould controversy. I will begin with some background about the aims of Morton's research. Then I will turn to the five most significant charges that Gould made against Morton. I will argue, contra Lewis et al., that most of Gould's arguments against Morton are sound. Although Gould made some errors and overstated his case in a number of places, he provided prima facia evidence, as yet unrefuted, that Morton mismeasured his skulls in ways that conformed to 19th century racial biases.

THE MORTON CRANIAL COLLECTION

Louis Agassiz (1807–1873), the Swiss-born Harvard Professor of Natural History, wrote to his mother: "Imagine a series of 600 skulls, mostly Indian, of all the tribes who now inhabit or formerly inhabited America. Nothing else like it exists elsewhere. This collection alone is worth a journey to America" (Kelly 1920, 824). Agassiz was referring to Morton's "American Golgotha," a collection of contemporary and ancient crania, which Morton spent much of life collecting. At the time of his death, this collection numbered over 1200 crania, and has since continued to grow to its present-day size of over 2000 crania. It is one of the largest and most representative collection of crania in North America, and perhaps the world.

Thanks to Morton's meticulous record keeping, much of which was published in his *Catalogue of Skulls of Man and the Inferior Animals* (1849a), we know quite a lot about the individual crania in his collection. For example, the bottom right skull in Figure 1 (labeled "804" by Morton) is a 12 year old girl of the "Pelasgic" form, found in the necropolis of Memphis. This skull was depicted in *Crania Aegyptiaca* on Plate 3 as Figure 2. Another item in the collection is #1346, the

> "Skull of a chief of the Lipan tribe of Indians, killed in a skirmish with Col. Doniphan's legion, on the 5th of May, 1847, at Poyo, near Parsos, in New Mexico. Man, aetat. 40. I. C. 84. This skull was procured and presented by Dr. A. Wislizenus, of St. Louis, Misouri." (Morton 1849, 55)

Fig. 1. The Morton cranial collection (photo credit Brett Calcott).

One of the most interesting specimens is #59, the skull of the infamous Australian cannibal Alexander Pearce: "Anglo-Saxon head: skull of Pierce, a convict and cannibal who was executed in New South Whales, A.D. 18–F.A. 85°. I.C. 99." In a footnote for this entry, Morton reproduces a letter from Wm. Cobb Hurry, Esq., of Calcutta, who received the skull from Mr. Crocket. Hurry reported that Pearce

> was a convict in Van Dieman's land, and escaped with others into the woods. Hunger compelled them to prey upon each other, till only Pierce and another were left. A romantic tale might be made from Pierce's own narrative of the feelings with which these two men watched each other, till overcome with fatigue, the last of the band fell a victim. Pierce was relieved by a party who fell in with him, and the cannibalism of which he had been guilty being attributed to necessity, was not punished. From that time his propensities acquired their full development and he succeed repeatedly in persuading his fellow prisoners to escape with him, for the sole purpose of killing them and devouring their flesh. He used to return secretly to the depot, and persuade a fresh victim that he had been sent by others who were waiting in the woods. He was at last caught; and being asked if he knew where one of his companions was, deliberately pulled an arm out of his jacket and showed it to the soldiers. Mr. Crockett, from whom I had this account, and who gave me the skull, is the Colonial Surgeon and attended Pierce in the hospital both before and subsequently to his crimes. He stated to me his conviction that Pierce was insane, which, however, did not prevent him from being hanged (1849, 5).

RACES.	No. of skulls.	Mean internal capacity in cubic inches.	Largest in the series.	Smallest in the series.
Caucasian.	52	87.	109.	75.
Mongolian.	10	83.	93.	69.
Malay.	18	81.	89.	64.
American.	147	80.	100.	60.
Ethiopian.	29	78.	94.	65.

Fig. 2. "On the Internal Capacity of the Cranium in the different Races of Men" from *Crania Americana*.

Why would a scientist spend so much of his time and fortune collecting skulls like these? The answer has to do with Morton's attempt to apply the methods of natural history to human variation. Although a physician by profession, Morton made substantial contributions to anatomy, geology, and natural history. He published works such as *Observations on the Geology and Organic Remains of the Secondary, Tertiary and Alluvial Formations of the Atlantic Coast of the United States of North America* (1828), *Synopsis of the Organic Remains of the Cretaceous Group of the United States* (1834), *Illustrations of Pulmonary Consumption* (1834), *Hybridity in Animals and Plants* (1847), and *An Illustrated System of Human Anatomy* (1849). When he set out to study humans, specifically the differences between human groups, he did so in much the same way as he approached the Cretaceous group. Specifically, he thought that by collecting a geographically diverse set of crania, he could understand variation among different races of humans.

Following Linnaeus and other naturalists of his period, Morton assumed that the best way to categorize human diversity was by using racial types. He relied on Johann Friedrich Blumenbach's (1752–1840) 5-fold categorization of human races: Caucasian, American, Malay, Mongolian, and Ethiopian (Morton 1839, 5). Morton further argued that each of these races had been "adapted from the beginning" (3) to their geographical locations, apparently rejecting the biblical account of human origins.

Morton's study of racial differences consisted of two stages. He began his investigations by reviewing the ethnographic literature,[1] such as it was, about different human races. Then he set out to give the ethnographic findings a physical basis in the properties of human crania. This involved both collecting the crania and also developing the techniques used to measure them. His review of the ethnographic literature, descriptions of his collection, and explanations of his measuring techniques are detailed in *Crania Americana* (1839), *Crania Aegyptiaca* (1844), and *Catalogue of Skulls of Man and the Inferior Animals* (1849a). These are the works which I will discuss throughout this article.

[1]His survey of the existing ethnographic literature fill Morton's notebooks. These notes can be found in "Samuel George Morton Papers," Series II: Notes. Library Company of Philadelphia.

MORTON ON RACE-TYPICAL TRAITS

As I discussed in the last section, Morton proposed to study and classify human beings using a combination of ethnographic and physical methods, and he describes his ultimate goal as developing an accurate classification of human beings (1839, 4). With his classification methods, he hoped to settle questions in genealogy (e.g., Are contemporary Native Americans related to ancient North, Central, and South Americans?) and history (e.g., Why were the Europeans able to invade North America so easily?).

Morton complied his ethnographic research in a long essay titled "On the Varieties of the Human Species," which is printed at the front of *Crania Americana*. To give some sense of the way that Morton combined physical anthropology with ideas about cognitive, personality, and behavioral traits, consider the following passages from this essay:

> The Caucasian Race is characterized by a naturally fair skin, susceptible of every tint; hair fine, long and curling, and of various colors. The skull is large and oval, and its anterior portion full and elevated. The face is small in proportion to the head, of an oval form, with well-proportioned features. The nasal bones are arched, the chin full, and the teeth vertical. This race is distinguished for the facility with which it attains the highest intellectual endowments (1839, 5).

Contrast this with what he says about Native Americans:

> The American Race is marked by a brown complexion, long, black, lank hair, and deficient beard. The eyes are black and deep set, the brow low, the cheek-bones high, the nose large and aquiline, the mouth large, and the lips tumid and compressed. The skull is small, wide between the parietal protuberances, prominent at the vertex, and flat on the occiput. In their mental character the Americans are averse to cultivation, and slow in acquiring knowledge; restless, revengeful, and fond of war, and wholly destitute of maritime adventure (1839, 6).

As one can see from these passages, Morton came to his study of the physical properties of crania with very strong assumptions about differences in lifestyle and intelligence between the races.

One of Morton's innovations in the classification of humans was to formally divide the five races into 22 families, and some of these into sub-families. For example, the Caucasian race is divided into seven families:

1. The Caucasian Family
2. The Germanic Family
3. The Celtic Family
4. The Arabian Family
5. The Libyan Family
6. The Nilotic Family
7. The Indostanic Family.

As an example of sub-families, the Arabian people are further divided into the Arabs, the Moors, the Bedouins, the Wahabys, and the Jews.

> The Jews or Hebrews were in their origin a pastoral nation, but in progress of time they established themselves in the cities of Palestine. Their physiognomy is familiar in the receding forehead, the elongated face, and the large and aquiline nose. Their high attainments in literature are fully attested by the sacred writings; and their zealous attachment to their religion, and their patient endurance of adversity, are among the most striking traits of their character. Dispersed by a divine judgment, they are to be found almost every where on the habitable earth, recognized by the same features, and the same undeviating form of worship (1839, 21).

Morton frequently commented on both physical characteristics as well as lifestyles. For example, about the Eskimos (Mongolian race, Polar family) he writes:

> The Eskimaux of Prince Regent's Bay, to the northeast of Baffin's Bay, and about 76° north, are of a dirty copper color, and very corpulent; while those on the west side of Baffin's Bay have clear complexions, which only become darker by old age and exposure.

He goes on to describe the lifestyle of a particular tribe from Greenland:

> They are crafty, sensual, ungrateful, obstinate and unfeeling, and much of their affection for their children may be traced to purely selfish motives. They devour the most disgusting aliments uncooked and uncleaned, and seem to have no ideas beyond providing for the present moment … With respect to the moral and intellectual character of this widely distributed family … they possess simplicity without silliness, and good sense without the art of reasoning. They are fickle and facetious, and their connubial infidelity is a proverb among voyagers (1839, 54–55).

Much of Morton's discussion of Caucasians illustrates the extent to which he thought they were culturally and intellectually superior to the other races. However, he also had plenty of racist and off colored things to say about some Caucasian sub-groups as well. Regarding the Russians:

> Under this ["selavonic"] denomination are also embraced the Russians, Poles, Lithuanians and part of the Bohemians and Hungarians. They are for the most part characterized by darker hair and complexion than the Teutonic tribes … The

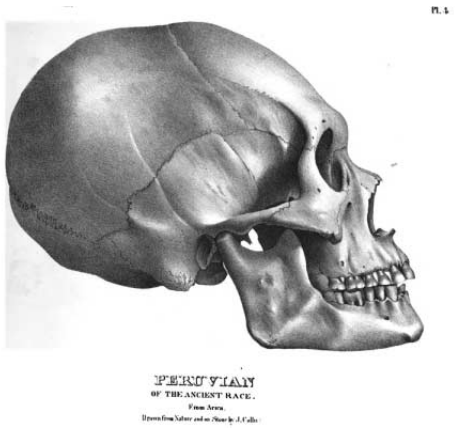

Fig. 3. John Collins' depiction of an Ancient Peruvian Cranium from Morton's collection.

> people of this division of the Germanic family are brave and enterprising, but generally rude and uncultivated; and the Russians, perhaps the most polished branch, emerged from the deepest barbarism so lately as the reign of Peter the Great (1839, 15).

Much of Morton's on essay on the "Varieties of the Human Species" reads in this way. But 95 pages into *Crania Americana*, Morton turns to his main subject: the physical measurements of the skulls themselves. For example, when discussing the cranium in Figure 3, an Ancient Peruvian (Morton's Plate 4 in *Crania Americana*), he writes:

> Our knowledge of [the Ancient Peruvians'] physical appearance is derived solely from their tombs. In stature they appear not to have been in any respect remarkable, nor to have differed from the cognate nations except in the conformation of the head, which is small, greatly elongated, narrow its whole length, with a very retreating forehead, and possessing more symmetry than is usual in skulls of the American race. The face projects, the upper jaw is thrust forward, and the teeth are inclined outward. The orbits of the eyes are large and rounded, the nasal bones salient, the zygomatic arches expanded; and there is a remarkable simplicity in the sutures that connect the bones of the cranium.
>
> The first idea that occurs to every one on looking at a series of these skulls is, that their peculiarities are in a great measure artificial. If, however, we carefully examine the cranium figured on the fourth plate, together with the accompanying smaller outlines, we find no evidence of mechanical compression. This head, on the contrary, appears to be of the natural form, unaltered by art; and it is figured as an illustrative type of the cranial peculiarities of the people now under consideration (1839, 97).

Such careful physical observations led Morton to some of his most interesting claims—ones that were controversial in his day, but are now accepted as correct. A large part of the argument in *Crania Americana* is that the ancient, highly sophisticated Native American cultures are part of a single group of which modern Native Americans are a part. Similarly, Morton argued in *Crania Aegyptiaca* that the ancient Egyptians and the modern Egyptians are the same people.

Although Morton deployed physical techniques, especially cranial measurements, to address questions about the unity and diversity of Americans and Egyptians, this is not the source of his fame. Of the 12 measurements he took for each skull, the most widely discussed by his supporters and detractors alike is the cranial capacity, the volume of the interior of the cranium, which is a proxy for the size of the brain. Morton reported the cranial capacity for each specimen in his collection, and, in each of his major craniological works, he gives summary tables of cranial capacity for each race, and for some racial families and sub-families.

In *Crania Americana*, the table is extremely simple. Morton merely reports the maximum, minimum, sample size, and mean for each of the five races. He writes, "Having subjected the skulls in my possession, and such also as I could obtain from my friends, to the internal capacity measurement already described, I have obtained the following results. The mean of the American Race (omitting the fraction) is repeated here merely to complete the table. The skulls of idiots and persons under age were of course rejected." Morton's table is shown in Figure 2.

Why was Morton so interested in cranial capacity? Although he does not say so directly, it seems clear that he regarded cranial capacity and intelligence to be tightly linked. *Crania Americana* contains a commissioned essay about the principles of phrenology written by "George Combe, Esq., the distinguished phrenologist." Combe gives the following argument to connect cranial capacity with intelligence (1839, 274):

> First. The brain of a child is small, and its mind is weak. As the brain grows in size and attains to maturity in structure, the mental manifestations increase in vigor. Secondly. A small brain is one but not the only cause of idiocy. A brain may be enlarged by disease and idiocy ensue; but if this organ be too small, although it be healthy in structure, idiocy is an invariable consequence … Thirdly. Individuals and nations distinguished for great aggregate force of mind, animal, moral and intellectual, have had large brains. King Robert Bruce, Napoleon, Cuvier, Canova, Burns the poet, Dr. Gall and Dr. Spurzheim, among men, and the Teutonic race compared with the Hindoo among nations, may be cited as examples (1839, 275–276).

Given this link between cranial capacity and intelligence, Morton believed that races could be ordered by average intelligence using physical methods. Further, this information

could be used to explain the racial differences that interested Morton.

The result of publishing tables of cranial capacity was quite predictable. Morton was taken to have established the innate differences in intelligence between races on empirical grounds. The Caucasian race, he believed, is "distinguished for the facility with which it attains the highest intellectual endowments" (1839, 5). Further, "Mongolians are ingenious, imitative, and highly susceptible of cultivation" (1839, 5), and while "the many nations which compose [the Ethiopian race] present a singular diversity of intellectual character … the far extreme is the lowest grade of humanity" (1839, 7). In other words, Morton took his measurements to have provided unbiased, empirical evidence for the claim that Caucasians are the most intelligent race, and Ethiopians (that is, black Africans) the least.

THE MISMEASURE OF MAN

Despite his fame in the 19th century, Samuel Morton faded into relative obscurity. Indeed, most modern scholars first learned of Morton from the work of Stephen Jay Gould. First in an article in *Science* (1978) and then 3 years later in his widely acclaimed book *The Mismeasure of Man* (1981), Gould brought Morton's work to light as a cautionary tale.

Gould's most famous accusation is that Morton literally mismeasured some of his collection, and that the measurements reported in *Crania Americana* were biased towards Caucasians. Specifically, Gould argued that where Morton made measurements errors, they tended to result in smaller African crania and larger Caucasian ones. Gould wrote: "I have reanalyzed Morton's [cranial measurements] and I find that they are a patchwork of assumption and finagling, controlled, probably unconsciously, by his conventional a priori ranking (his folks on top, slaves on the bottom)" (1978, 504). Far from being the greatest empiricist of his age, Morton would be better seen as a prime example of how unconscious bias permeates the scientific enterprise.

The reception of Gould's book was mixed, although mostly positive. The New York Times book review said that the book offered a "a sharp blow … to the apostles of innate, hereditary, hierarchical intelligence in human beings." (Lehmann-Haupt 1981) The London Review of Books said that "Stephen Gould's contribution to this last debate is to open one or two coffins containing the scientific skeletons of the past with the purpose of nailing down the lids even more securely. From the point of view of the modern debate, they were better left undisturbed, but *The Mismeasure of Man* is fine history, and relevant to the present-day controversy insofar as it explains its social background …" (Edwards 1982). Others, however, were not so impressed. In reviewing the second edition of *Mismeasure*, the controversial race researcher J. Philippe Rushton accused Gould of "several counts of scholarly malfeasance" (Rushton 1997).

Despite criticisms in some quarters, Gould's work was widely discussed and widely taught. It was used to illustrate examples of racial bias in science, and its discussion of the origins of IQ testing were highly influential among those who opposed the scientific findings and policy recommendations of researchers like Arthur Jensen, Richard Herrnstein, and Philippe Rushton. Although there was debate about these aspects of Gould's arguments, there seemed to be little attention paid to Gould's detailed discussions of Morton.

The first reexamination of Gould's discussion of Morton came in John S. Michael's undergraduate honors thesis, which was later published in *Current Anthropology*. Michael remeasured 201 crania from Morton's collection and determined that "Individual cranial capacities thus determined were consistent with Morton's but on the average 32.48 cm^3 (roughly 2 in^3) lower" (Michael 1988, 351). He thus argued that

> Contrary to Gould's interpretation, I conclude that Morton's research was conducted with integrity. Morton was one of the first scholars to attempt the study of human diversity through objective measurements, and it is not surprising that he made mistakes. Although he cannot be excused for his errors or his unfair comparison of means, he should be given credit for having taken the risk of experimenting with a new and innovative technique (353).

Michael's publication never received much scholarly attention, and the attention that it did receive was often from researchers who were writing critical pieces about Gould. And while the Morton Cranial Collection itself remained an important research tool, Morton's works had fallen into relative obscurity, except insofar as they were seen to be the object of Gould's attacks in *Mismeasure*. This all changed in the spring of 2011.

REMEASURING MAN

In 2011, a team of researchers affiliated with the University of Pennsylvania Museum of Archeology and Anthropology published a response to Gould's accusations of bias. Unlike Gould who based his conclusions on an analysis of Morton's meticulously documented books and journal articles, the Penn team worked with Morton's collection itself. They remeasured a sample of 308 skulls using precision plastic balls and reanalyzed Gould's tabulations and statistics. Their conclusion: "Ironically, Gould's own analysis of Morton is likely the stronger example of bias influencing results." (Lewis et al. 2011, 5) They accuse Gould of an egalitarian bias, claiming that he skewed his analysis and interpretation to conclude that there are no racial differences in cranial capacity in Morton's collection. Their own main conclusions are that Morton's measurements were mostly accurate and where they were inaccurate, they show no evidence of racial bias. In the following sections I will discuss and assess

Gould's five most important criticisms of Morton as well as Lewis et al.'s replies.

Did Morton mismeasure crania to conform to his racial biases?

In his *Science* article and again in *Mismeasure of Man*, Gould accuses Morton of mismeasuring crania to conform to his racial biases. Specifically, Gould argues that the average measurement Morton reports for Africans ("Ethiopians") in *Crania Americana* was abnormally small. In order to understand Gould's argument, we need to understand Morton's method for measuring cranial capacity.

All of Morton's cranial capacity measurements were carried out by filling the cranial cavity with spherical material, and then measuring the volume of this material. To make his measurements of cranial capacity for *Crania Americana*, Morton used white "pepper seeds."[2] Although Morton took precautions against the seeds varying in size by sifting them, he later acknowledged that the seed-based measurements led to unreasonably high variances in the measurements. This was probably caused by uneven settling of different-sized seeds. Concerns about these inaccuracies ultimately led Morton to use BB shot instead of seed for the measurements given in *Crania Aegyptiaca* and his *Catalogue*.

For the purpose of his argument, Gould accepts Morton's later measurements with BB shot as precise and accurate (he calls them "objective"). He then compares the mean cranial capacities from seed-based measurement to the mean cranial capacities of shot-based measurements. This comparison is not straightforward, because Morton only provided individual measurements for Native American skulls in *Crania Americana*; all the rest were averaged together by race.[3] Thus, Gould had to infer which African, American, and Caucasian crania listed in the *Catalogue* were measured with seed.

After working out the relevant subsamples of Africans, Americans, and Caucasians, Gould points out that, upon remeasurement for *Crania Aegyptiaca*, the means of all three samples had changed, but not by the same magnitude. This differential change across the three racial means is illustrated in Figure 4.

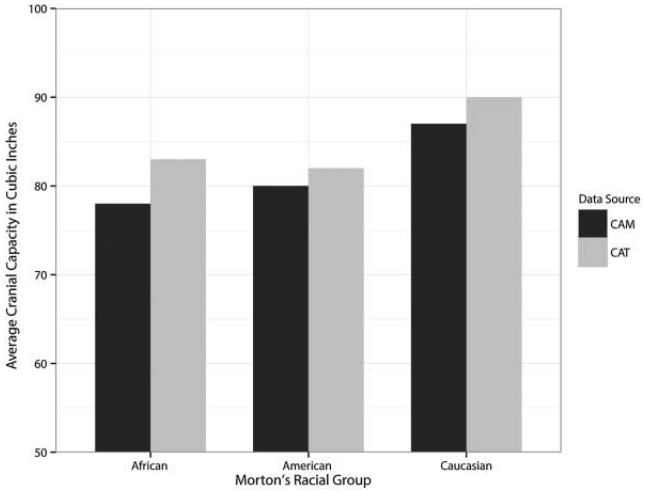

Fig. 4. Change in mean cranial capacity from *Crania Americana* (1839) to *Catalogue of Skulls of Man and the Inferior Animals* (1849).

In Figure 4, we can see that all of the 1849 means were larger, presumably because shot packs better than seed in virtue of its uniformity and its weight. However, we also see that the magnitude of change is different for the three groups. The African skulls have a much larger change in mean cranial capacity than the Americans and Caucasians. If this difference was the result either of lack of precision or of a systematic measurement error, the change should be approximately the same for each race. Gould thinks that the best explanation of this is unconscious manipulation on Morton's part (or, perhaps, on the part of the assistant who did the measurements).

As a potential explanation of the discrepancy, Gould writes: "Plausible scenarios are easy to construct. Morton, measuring by seed, picks up a threateningly large black skull, fills it lightly and gives it a few desultory shakes. Next, he takes a distressingly small Caucasian skull, shakes hard, and pushes mightily at the foramen magnum with his thumb. It is easily done, without conscious motivation; expectation is a powerful guide to action…" (1981, 65). Whether by this or some other mechanism, Gould concluded that Morton let his racial biases affect his measurement when he was using a technique that was less accurate and potentially open to bias.

In contrast to Gould's analysis of Morton, Lewis et al. argue, on the basis of their own remeasurement, that there is no evidence of racial bias. They performed the heroic task of remeasuring 46% of the crania whose capacity was reported by Morton. They published their raw measurements, the differences between their measurements and Morton's, and a statistical analysis of their data. Their analysis first calculates a "current shot adjusted" measurement for each skull by adding a correction factor for change of technique (shot apparently overestimated by an average of $3\,in^3$). They subsequently performed linear and quantile–quantile regressions to determine

[2]In *Crania Americana*, Morton describes his measuring material as "pepper seeds." Gould and Lewis et al. assume he meant mustard seeds. Based on his descriptions of the seeds, I believe that Morton is actually talking about "white peppercorns," the common modern American name for the seeds of the of Piper nigrum plant whose skin has been removed. In an 1849 journal article (Morton 1849b) and in Morton's *Catalogue*, he refers to the pepper seeds as "white mustard seeds," making the correct identification of this material very difficult.

[3]I have relied on Lewis et al.'s reconstruction of the sample as listed in their Dataset S3 from Lewis et al. (2011).

if any of Morton's shot-based measurements differed significantly from their own after correction. They determined that just 2% of Morton's measurements differed significantly.

Their next step was to perform a binomial test to determine whether or not Morton's overestimates were centered on his Caucasian skulls and his underestimates were centered on his non-Caucasian skulls. The only significant finding in this test was that Morton overestimated more Egyptian crania then would have been predicted by chance. However, these overestimated crania were Morton's Negro Egyptians, so this is exactly the opposite of what we might predict had Morton's errors been guided by unconscious racist bias. Lewis et al. conclude that "These results falsify the claim that Morton physically mismeasured crania based on his a priori biases" (3). In other words, Lewis et al. claim that there is no evidence of mismeasurement in Morton's published data, and certainly no evidence of racial bias.

Despite what they say, Lewis et al. have not falsified Gould's claim. Gould's argument was that the differences between Morton's seed-based and shot-based measurements reveal Morton's racially biased measurements. Showing that Morton's shot-based measurements are reliable is very impressive, but cannot support the claim that Morton's work reveals racial bias. Indeed, Gould accepts the reliability of Morton's shot-based measurements, since he has no other way of determining differences with the seed-based measurements.[4]

In a later section of the article, Lewis et al. seem to acknowledge this problem implicitly because they give a second argument against Gould. Since Gould's argument depends on comparing seed-based and shot-based measurements, he needed to know which skulls were actually measured with seed. This is hard to determine because the collection grew between 1839 and 1849, and Morton only reports individual seed-based measurements for Americans. Because of this, Gould had to make educated guesses about which skulls Morton measured for his 1839 book. This, Lewis et al. argue, renders Gould's analysis "highly questionable" (p. 4).

If this was truly questionable, then Lewis et al. would fall prey to the same critique they level against Gould, since they also had to make educated guesses about which crania were measured in 1839. Although uncertainty is introduced by these guesses, they make convincing arguments based on inferences from catalog number to the probable date of acquisition. If Lewis et al. can do this successfully, then so could Gould. All parties had the same data available.

[4]Lewis, et al. also compare their remeasurements of Native American crania to some of the specific seed-based cranial capacities reported in *Crania Americana*. Although they show the relative reliability of these measurements, this also cannot address Gould's argument, which was based on the differential reliability of Morton's seed-based measurements among racial groups, especially Africans.

Lewis et al. also point out another respect in which they think Gould's comparison of the earlier and later measurements is deficient. Although Gould is correct to claim that the average cranial capacity for Americans changes 2.2 in^3, they argue that this doesn't give the complete picture. The range of change for individual measurements varies from $+12\ in^3$ to $-10\ in^3$ with a standard deviation of 2.8 in^3. They also argue that the changes do not appear to be patterned. One Peruvian cranium increased by 18%, while another decreased by 7%. "This," Lewis et al. write, "casts significant doubt on the hypothesis that mismeasurements with [seed] were a function of Morton's racial bias." (p. 4) They draw this inference, it seems, because not all of the Americans were under-measured using seed.

This second argument does not improve Lewis et al.'s case against Gould. First, the fact that Morton doesn't report the specific skulls he measured in *Crania Americana* is a red herring; both Gould and Lewis et al. were able to make a reasonable guess about which skulls Morton was working with by using Morton's reported means and acquisition dates, and working backward to construct the relevant sample. This involved some amount of inference and assumption, but to call it "highly questionable" is an overstatement.

Moreover, Lewis et al.'s report about the range and standard deviation is an important point, but does not diminish Gould's argument. If we had all of Morton's individual measurements from *Crania Americana*, we could tell a more complete statistical story. Unfortunately, these data have not yet been located and may no longer exist. But since neither Gould nor Lewis et al. have these data, Gould did nothing illegitimate in comparing the means. Even if there is high variance in the mean, systematic differences in means between sub-samples when techniques change is evidence of systematic measurement error. Gould's argument relying on these measurements is entirely legitimate and provides him with prima facia evidence of a systematic error. This error could be a sign that Morton's racial bias led him to unconsciously mismeasure cranial capacity with seed for *Crania Americana*.

Lewis et al. point out correctly that an alternative explanation consistent with this pattern is that Morton's assistant, who was taken off the job, was simply poor at taking measurements. But unless there was some temporal ordering to the measurements, meaning that he got better or worse as time went on, it is very hard to account for the systematic differences other than as an instance of bias on the part of Morton or his assistant.

Did Morton manipulate his inclusion of subsamples to depress the grand means?

Gould alleges that Morton had a useful tool for unconsciously manipulating the grand means to match his a priori expectations: He could include or exclude sub-samples of different sizes. "As a favorite tool for adjustment, Morton chose to include or delete large subsamples in order to match grand means with a priori expectations" (Gould 1978, 508).

For example, Morton could make the Indian sample mean smaller by including more Peruvians, which were known by Morton to have smaller heads on average. Between 1839 and 1849, Morton's sample of Peruvians had indeed changed. In *Crania Americana* 23% of the American sample were Peruvians, but by the time he tabulated his final summary 1849, 50% of the sample were Peruvians. This led to a decrease in Morton's mean cranial capacity of Americans. Or as Gould dramatically described the change: "Morton's Indian mean had plummeted to 79.3 in^3 [in 1849]" (508).

Lewis et al. respond to Gould in two ways. First, they point out that the "plummet" between 1839 and 1849 was from 79.6 to 79.3 in^3, so this isn't really much of a plummet. Moreover, they point out that Morton's 1849 figure for the average cranial capacity of Americans was computed by equally weighting averages for the Peruvians, Mexicans, and Barbarous tribes. This being the case, simply including more Peruvians in the overall sample couldn't change things very much, unless he was able to find much smaller-headed Peruvians. And when we look at the details, we find that the Peruvian mean cranial capacity changed by less than 1 in^3 from 1839 ($n = 33$) to 1849 ($n = 155$).

At least with respect to the Peruvians, it seems clear that Gould's accusation that Morton manipulated subsamples, even unconsciously, does not stand up to scrutiny. Lewis et al. are clearly correct that a 0.3 in^3 drop is not a very large change. Moreover, it is also clear that there wasn't much latitude for manipulation of subsample size when Morton had already adopted the tallying method of reporting weighted means for his subpopulations.

While Gould's accusations against Morton fail here, there is a subtlety to Gould's argument that hasn't yet been brought out and that remains correct. Although Lewis et al. focus on Gould's discussion of sample size, Gould also discusses the problematic ways that Morton determines which samples to include in racial means. For example, Morton included small-headed Inca Peruvians in the American mean, but excluded small-headed Hindus from the Caucasian mean. This allowed Morton's American sample to appear smaller and Caucasian sample to appear bigger than it might have been otherwise.

Gould's criticism can be made even more generally. Morton already saw differences in cranial capacity among the families and sub-families in his samples. These differences should have led him to be much more cautious about reporting means for each race, since there is much variation within races, and relatively small differences between the racial aggregates. Thus while Lewis et al. are correct in their specific critiques of Gould, Gould's more general critique is important and correct.

Was sexual dimorphism an additional source of bias in Morton's measurements?

Gould argues that there is a very striking main trend that jumps out of Morton's data: sexual dimorphism in size. Female skulls

Ethnographic Division.	Locality.	No. of Crania.	Largest Brain.	Smallest Brain.	Mean.	
Pelasgic Form.	Memphis.	14	97	79	89	Mean, 88 C. I.
	Abydos.	1	89	89	89	
	Thebes.	5	92	82	86	
	Philæ.	1	74	74	74	
Semitic Form.	Memphis.	1	88	88	88	Mean, 82.
	Abydos.	1	69	69	69	
	Thebes.	3	85	79	79	
Egyptian Form.	Memphis.	7	83	73	79	Mean, 80.
	Abydos.	2	96	85	90	
	Thebes.	25	95	68	80	
	Ombos.	2	77	68	73	
	Debôd.	3	82	70	75	
Negroid Form.	Maabdeh.	1	71	71	71	Mean, 79.
	Thebes.	5	88	71	81	
Negro.	Philæ.	1	73	73	73	Mean, 73.

Fig. 5. Morton's summary table from *Crania Aegyptiaca*.

are almost always smaller than male skulls. Of course, Morton could only be sure of sex from contemporary skulls or ones that came along with preserved remains.[5] For others, he had to infer sex by skull size. However he did it and however reliably he did it, by the time Morton published his 1849 catalogue; he had labeled most of his collection by sex.

Despite labeling the crania by sex, in his summary tables, Morton never reports different averages for males and females. The example Gould examines in detail compares the Caucasians to the Negroids in *Crania Aegyptiaca*.

If we start from the summary table shown in Figure 5 and average the Pelasgic, Semitic, and Egyptian caucasians, we get a weighted Caucasian mean of 83 in^3. We can see directly from the table that the Negroids have a mean of 79 in^3.

In the text, Morton identifies 24 of these Caucasians as male and 22 Caucasians as female. For the males, the average cranial capacity is 86 in^3 (SD 6.6) and for females it is 77 in^3 (SD 6.4), a difference of 9 in^3. Morton identified two of the Negroids as females (71 in^3 and 77 in^3) and didn't determine the sex for four others. The four unidentified crania had cranial capacities of 77,

[5]There is some disagreement about how many crania came along with remains. Gould claims that many of the ancient Egyptians were either mummified skulls or came along with remains (1996, p. 94). In personal communication, Janet Monge, keeper of the Morton collection at Penn Museum and a co-author of Lewis et al., denies this, although the dispute seems to be quantitative.

77, 87, and 88 in^3. In his 1849 catalogue, he identifies the sex of these four crania, presumably by size. Of these four, he identifies one as female, the other three as male. This gives a male Negroid mean of 84 in^3 and female Negroid mean of 75 in^3. Gould notes that Morton's sexual identities are suspicious: why should he have identified one of the 77 in^3 crania as female, and the other as male? Had he identified them both as female, then the male mean would be 88 in^3 and female mean would be 76 in^3.

With these better estimates of cranial capacity by sex, we can start comparing like to like. If we compare the Negroid male mean to the Caucasian male mean, we see the Negroid males actually have higher mean cranial capacity to the Caucasian males. The opposite is true for Caucasian and Negroid females.

From this analysis, Gould concludes that the apparent difference in cranial capacity between Caucasians and Negroids in *Crania Aegyptiaca* only reflects the fact that about half the Caucasian sample is male, while only one-third of the Negroid sample was male. When one compares males to males and females to females, one sees that there is little difference, and what difference there is may actually favor the Negroids. More generally, Gould argues that this is a demonstration of how sexual dimorphism in size may be driving many of Morton's results.

In discussing this aspect of Gould's argument, Lewis et al. write, "Certainly, more accurate population averages would be obtained if each sample were composed of equal numbers of males and females … [but it] is essentially impossible for Morton to have exploited sexual differences … [because] Morton did not collect the skulls himself, and there is no evidence that he excluded any skulls from measurement on the basis of sex. Indeed, Morton was largely blind to the sex of the skull … because of the low accuracy of determining sex from the skull …" (Lewis et al. 2011, 5). In other words, because Morton did not and could not have manipulated the number of males and females in his collection, this could not have been a source of bias.

This is a weak argument. Gould did not accuse Morton of consciously or even unconsciously manipulating the means on the basis of sexual dimorphism. He said that the dimorphism may have been an additional source of bias in Morton's results; more males will result in a larger mean cranial capacity. Moreover, this bias should have been obvious to Morton, but he overlooked it or chose to ignore it in computing his summary tables. Understood this way, Gould's argument is unambiguously correct.

Are Native Americans really the largest Crania in the collection

One of Gould's most interesting claims is that Native Americans, not Caucasians, have the largest crania in Morton's collection. Morton's *Catalogue* reports that Americans have a mean cranial capacity of 79 in^3, while Caucasians have a mean crania capacity of 85 in^3. Gould's reanalysis of Morton's data yields an American mean of 86 in^3, giving them the largest cranial capacity.

Table 6. Corrected values for Morton's final tabulation.

People	Cranial capacity (in^3)
Native Americans	86
Mongolians	85
Modern Caucasians	85
Malays	85
Ancient Caucasians	84
Africans	83

Fig. 6. Table 6 from Gould (1978).

How did Gould come to this figure for the American skulls? In Figure 6, I reproduce Table 6 from Gould (1978), which summarizes his results and which shows Native Americans as having the largest cranial capacity. Gould arrived at his new value for the Native American mean cranial capacity by adopting the following three restrictions:

1. Weight samples for all sub-populations evenly.
2. Only include crania for sub-populations with more than three samples.
3. Only include crania that were in Morton's collection at the time of *Crania Americana* and so had originally been measured with seed.

The justification for weighting sub-samples evenly should be obvious, and indeed was followed by Morton in the tables for his catalogue. Gould imposed the second restriction because of the aforementioned issue of sexual dimorphism. He explains this restriction as follows: "I chose four skulls as a minimum subsample size in order to give a probability near 5% that the group would represent a single sex only … If (as may well not have been the case) Morton's skulls had an equal chance of being male or female, then only 1 in 16 samples of four skulls would form a unisexual group" (1978, 509, n. 22). Finally, Gould gives no justification for the third restriction and indeed the only way to determine that he used it, other than re-tallying the mean, is to note that he said that he "recomputed the means for skulls remeasured with shot …" (1978, 508, my emphasis). Using these restrictions and examining the measurements that conform to them, Gould concludes that Americans, not Caucasians, have the highest cranial capacity.

Lewis et al. reply to Gould's argument by noting that restriction 3 is totally arbitrary. "… Morton's publications and analyses for his seed- and shot-based measurements are completely separate …" (Lewis et al. 2011, 4). They further point out that Gould did not apply this restriction to any other sample, making this restriction even more arbitrary. When this restriction is lifted, but leaving the other two in place, Lewis et al.

claim that the correct value for Native American cranial capacity is 83 in^3.[6]

It is clear that Lewis et al. are correct here. Gould's arbitrary restriction to the *Crania Americana* sample resulted in an overly large cranial capacity for the Americans. While it was important to compare the same "seed sample" measured with both seed and shot for changing means argument discussed in Section A, Gould should have used Morton's most complete sample to argue that there are "no differences between Morton's races" because "all have means between 83 and 86."

Moreover, this is an especially unfortunate mistake because the correct number makes Gould's point just as well as the inflated number. Gould wanted to argue that all the averages are within a small range of values (mean $\Delta 3$ in^3). Using the correct number, one can still argue that Morton's calculated average of 79 in^3 is not reliable. Lewis et al.'s estimate of 83 in^3 for the American cranial capacity will place the Americans at the lower end of the cranial capacity scale, but it is now a very small range of differences. So while Morton's American crania do not have the largest cranial capacity, analyzed this way, they are the same size as the Africans, and only slightly smaller than the Caucasians. This was the point that Gould wanted to establish, and he should have done it without the arbitrary restriction.

Did Morton's errors really conform to his bias?

Gould's overarching critique of Morton was that, where there were errors in his measurements and analyses, these errors corresponded to Morton's racial biases. Lewis et al. argue that, in a twist of irony, Gould's preferred method of computing means conforms more to Morton's biases than Morton's own means.

For the sake of argument, Lewis et al. grant Gould his figure of 86 in^3 for the Americans. They then go on to claim that Gould made arithmetic errors in his Science article, miscomputing the Modern Caucasian mean[7] and that the correct number is 87 in^3. This moves the Modern Caucasians from mid-pack to the top.

Lewis et al. point out that, with this corrected table (Table 1), the ranking of cranial capacity is Caucasians > Native Americans > Mongolians and Malays > Africans. In contrast, Morton's table had Caucasians > Malays > Africans > Mongolians > Native Americans.

Although he generally had a low opinion of Native Americans, Morton seemed to regard these peoples as more intelligent than the other non-Caucasian races. Given this, Lewis et al.' s "Corrected Table 6" (Table 1) seems to correspond more closely to how Morton would have ranked the races than his own table. In contrast, Morton's own table puts Native Americans at the bottom, below Africans, of whom he thought especially poorly (2011, 4).

Table 1. Lewis et al.'s corrected Table 6

People	Cranial capacity (in^3)
Modern Caucasians	87
Native Americans	86
Mongolians	85
Malays	85
Ancient Caucasians	84
Africans	83

This twist would indeed be ironic, but it is based on a false premise; Gould's original computed mean (reported in Gould 1978) for the Caucasians is correct. Lewis et al. do not say how they computed their figure of 87 in^3, but it appears to come either from following Gould's erroneous computation in *Mismeasure* or from an invalid method of tabulating weighted means. A closer examination of Morton's 1849 table shows how this error may have come about.

Figure 7 contains part of Morton's summary table from his catalogue. If we average the values for the Teutonic, Pelasgic, Celtic, Indostanic, Semetic, and Nilotic families, we get Gould's original figure of 85 in^3.[8] However, if we average the values for

TABLE,

Showing the Size of the Brain in cubic inches, as obtained from the measurement of 623 Crania of various Races and Families of Man.

RACES AND FAMILIES.	No. of Skulls.	Largest I. C.	Smallest I. C.	Mean.	Mean.
MODERN CAUCASIAN GROUP.					
TEUTONIC FAMILY.					
Germans,	18	114	70	90	92
English,	5	105	91	96	
Anglo-Americans,	7	97	82	90	
PELASGIC FAMILY.					
Persians, Armenians, Circassians,	10	94	75	84	
CELTIC FAMILY.					
Native Irish,	6	97	78	87	
INDOSTANIC FAMILY.					
Bengalees, &c.	32	91	67	80	
SEMITIC FAMILY.					
Arabs,	3	98	84	89	
NILOTIC FAMILY.					
Fellahs,	17	96	66	80	
. . . . .					

Fig. 7. Portion of Morton's 1849 table of racial families.

[6]Lewis et al. note that there is some variance in this number depending on the exact grouping scheme and that Morton is not always consistent. I have used the average which includes all groups. The highest value is obtained by eliminating the Araucanian and Otoe/Winnebago samples, which increases the mean to 84.

[7]In his 1978 *Science* article, Gould says the number is 85.3 in^3. In *The Mismeasure of Man*, Gould give the value of 87 in^3 and mentions in a footnote that he thinks the original value was the result of his misreading a photocopy of Morton's table. As I will show below, Gould was correct originally and his revised value, endorsed by Lewis et al., is incorrect.

[8]We might also apply Gould's own restriction of not including any group with $n \leq 3$. If we did this, we would exclude the Arabs. This doesn't change the weighted value within the two significant figures precision I am adopting throughout. The value remains 85 in^3.

Table 2. Author's corrected Table 6

Morton's racial category	Morton's reported 1849 average cranial capacities (in^3)	Reconstruction of the 1849 cranial capacity table (in^3)
Modern Caucasians	85	85
Mongolians	82	84
Malays	85	85
Americans	79	83
Africans	83	83

the Pelasgic, Celtic, Indostanic, Semetic, and Nilotic families and also the German, English, and Anglo-American subfamilies, then we get Lewis et al.'s figure of $87\,in^3$. But this is a computational error. Either we should equally weigh all the families or all the sub-families, but we shouldn't weigh a mixture of families and sub-families. Morton himself was careful to avoid this error, and he even reports a family mean for the Teutonic family, which he then averages with the other family means.

When we keep Gould's original figure for Caucasians and accept Lewis et al.'s amended value for the Native Americans based on lifting Gould's third restriction, we can generate Table 2, explained in detail in Appendix A.

The second column of Table 2 reflects my best reconstruction of Morton's average measurements for the different racial categories. Importantly, this table confirms Gould's contention that "There are no differences to speak of among Morton's races." (1978, 508). The total difference between the racial categories is reduced to $2\,in^3$, considerably smaller than the variation in these groups. So even if Morton's original numbers didn't conform to racial stereotypes in quite the way Gould thought, Gould's overall and most important point is correct. There is little difference in mean cranial capacity between the races, and Morton did not see this. Moreover, there is far less difference between the races than between the sexes.

MORTON VERSUS GOULD: AN ASSESSMENT

> There have been dark mutterings for a decade or more that *Mismeasure* was PC revisionism. (Kim Sterelny)

Although I haven't discussed every single argument in Gould and in Lewis et al., I have covered their major points of contention about measurement, sexual dimorphism, groupings, and the proper values for the infamous Table 6. Although it should be clear from my discussion above that Gould made some false moves, we should ask if he was mostly correct, or if his book was an example of PC revisionism.

Before answering this question, I think it is important to state explicitly that Gould's Science article and his book were politically motivated. He is extremely open about this in the introduction to the second edition of *Mismeasure*.

> My original reasons for writing *The Mismeasure of Man* mixed the personal with the professional. I confess, first of all, to strong feelings on this particular issue. I grew up in a family with a tradition of participation in campaigns for social justice, and I was active, as a student, in the civil rights movement at a time of great excitement and success in the early 1960s. (Gould 1996, 36)

The issue is raised again when he discusses differences in intellectual style with one of his arch nemeses, American Enterprise Institute fellow Charles Murray:

> … I have never been able to understand why [Murray] insists on promulgating the disingenuous argument that he has no personal stake or preference in the subject of The Bell Curve, but only took up the study from disinterested personal curiosity … He has been employed by right-wing think tanks for years, and they don't hire flaming liberals … If I were he, I would say something like: "Look, I'm a political conservative, and I'm proud of it. I know that the argument of The Bell Curve meshes well with my politics. I recognized this from the beginning. In fact, this recognition led me to be especially vigilant and careful when I analyzed the data of my book." (Gould 1996, 37–38)

My second preliminary point is that despite the way some scholars have understood his arguments (e.g., Kitcher 2004), Gould never claims that he has remeasured anything. In fact, he says that his relevant expertise lies in data analysis, especially factor analysis.

> The subject that I did choose for *The Mismeasure of Man* represents a central area of my professional expertise … the statistical analysis of genetically based variation within and between populations … I was trained primarily in the granddaddy of multivariate techniques … [especially] factor analysis. (Gould 1996, 40)

Gould claims that he brings expertise in statistics, not history, psychology, or physical anthropology to the table.

Applying his expertise in statistics to Morton's measurements and analysis, Gould concludes that Morton literally mismeasured skulls with seed to make the Africans look smaller and "his people" look bigger. Despite what Lewis et al. claim, this argument has not been called into question by their remeasurements. They rightly point out that another possible explanation, indeed Morton's own explanation, is that Morton's assistant was poor at taking cranial capacity measurements. This is entirely possible and would have had the effect of making the measurements performed with seed generally inaccurate. But

Gould's empirical point remains. There is prima facie evidence of racial bias in Morton's (or his assistant's) seed-based measurements. This argument is based on Gould's accurate analysis of the difference between the seed- and shot-based measurements of the same crania.

Gould is also correct about two other major issues. First, sexual dimorphism is a very suspicious source of bias in Morton's reported averages. Since Morton identified most of his sample by sex, this is something that he could have investigated and corrected for. Second, when one takes appropriately weighted grand means of Morton's data, and excludes obvious sources of bias including sexual dimorphism, then the average cranial capacity of the five racial groups in Morton's collection is very similar. This was probably the point that Gould cared most about. It has been reinforced by my analysis.

However, Gould's analysis is flawed in a number of important ways. Although he doesn't say this outright, Gould implies that there are no patterns of differences in the collection's cranial capacity. This is incorrect, because, among other things, cranial capacity varies with overall body size, and overall body size varies with the distance from the equator. Anthropologists call this fact Bergmann's rule (Bergmann 1847) and it is true for many mammals.

This and other types of geographically patterned variation can be found in Morton's crania. In an exhibit about Morton's collection titled **Making and Unmaking Race**, my fellow curators and I write: "When measuring crania many different patterns emerge. For instance, the crania in this case show how head shape slowly becomes more rounded from Southern Africa to Northern Europe. On the other hand, cranial capacity gets larger as the climate gets cooler north and south of the Equator. These two traits, like so many others, change independently, and do not align to racial categories" (Monge et al. 2012). To be fair, Gould never says that all crania are the same. But a more nuanced look at how crania and other physical traits vary strengthens Gould's argument that there are no general differences among different races as defined by Morton and his contemporaries.

Finally, Gould makes a substantial mistake in his exclusion of all Native Americans that weren't in Morton's 1839 sample. There is no principled reason to do this. Moreover, even without this exclusion, Gould's point is preserved. Native Americans are no longer abnormally low in mean cranial capacity.

In the final analysis, Gould's charge that Morton's analyses exhibit racial bias seems well-justified. Gould made some analytical errors which were uncovered by Lewis et al., but his two most important claims—that there is evidence that Morton's seed-based measurements exhibit racial bias and that there are no significant differences in mean cranial capacities across races in Morton's collection—are sound. Despite the continuing importance of Morton's collection for modern research, albeit research that rejects Morton's assumptions about racial categories and the link between cranial capacity and intelligence, his work remains a cautionary example of racial bias in the science of human differences. Gould's analysis and critique of Morton has been widely discussed and widely used as an illustration of implicit bias in science; it ought to remain so.

Acknowledgments

Many thanks to Peter Godfrey-Smith, Philip Kitcher, Janet Monge, Kim Sterelny, and Michael Yudell, and especially Paul Mitchell for helpful comments and discussions about earlier drafts of this manuscript.

REFERENCES

Bergmann, C. 1847. Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. *Göttinger Studien* 3: 595–708.

Edwards, A. W. F. 1982. When Gould meets Galton. *London Rev. Books* 4: 13.

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

Gould, S. J. 1981. *The Mismeasure of Man*. WW Norton & Company, New York.

Gould, S. J. 1996. *The Mismeasure of Man*, 2nd Ed. WW Norton & Company, New York.

Kelly, H. A. 1920. *A Cyclopedia of American Medical Biography: Comprising the Lives of Eminent Deceased Physicians and Surgeons From 1610 to 1910*. W.B. Saunders Company, Philadelphia.

Kitcher, P. 2004. Evolutionary theory and the social uses of biology. *Biol. Phil.* 19: 1–15.

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

Monge, J., Cox, S., Weisberg, M., and Yudell, M. 2012. *Year of Proof: Making and Unmaking of Race*. University of Pennsylvania Museum of Archeology and Anthropology, Philadelphia.

Morton, S. G. 1839. *Crania Americana or, a Comparative View of the Skulls of Various Aboriginal Nations of North and South America*. John Pennington, Philadelphia.

Morton, S. G. 1844. *Crania Aegyptiaca. Or, Observations on Egyptian Ethnography, Derived from Anatomy, History, and the Monuments*. John Pennington, Philadelphia.

Morton, S. G. 1849a. *Catalogue of Skulls of Man and the Inferior Animals*. Merrihew and Thompson, Philadelphia.

Morton, S. G. 1849b. Observations on the size of the brain in various races and families of man. *Proc. Acad. Nat. Sci. Philadelphia* 4: 221.

Morton, S. G. (ca 1839 –1849). Papers, Series II: Notes. Library Company of Philadelphia.

Trivers, R. 2012. Fraud in the Imputation of Fraud. Psychology Today, October 4, 2012. http://www.psychologytoday.com/blog/the-folly-fools/201210/fraud-in-the-imputation-fraud.

Appendix A: Explanation of tabulations and calculations.

Figure 4 The bar graph in Figure 4 depicts the mean cranial capacity reported in *Crania Americana* and the *Catalogue of Skulls of Man* and the Inferior Animals. The graph is based on the values in the table below.

Morton's racial category	1839	1849	Change
Caucasian	87	90	+3.0
American	80	82	+2.0
African	78	83	+5.0

The column labeled 1839 contains the average cranial capacity measurements reported by Morton in *Crania America*. Although the printed version of the book listed 82 as the mean cranial capacity for Americans, most of the existing copies that have been examined contain a hand correction to 80 cubic inches. This is the correct average based on the data Morton reports in that book.

Column 2 contains the cranial capacity values for the crania remeasured with shot reported in Morton's *Catalogue*. I have followed Lewis et al.'s reconstruction of the sample, which they, in turn, reconstructed from Gould's reported means. These samples have not been weighted by sub-group in order to make a direct comparison between the measurements reported in *Crania Americana* and those reported in the *Catalogue*. Note that Gould and Lewis et al. use four significant figures, when in many cases Morton only reports 2. I thus report two significant figures in this table.

Calculation of Table 2

The table below is labeled "Table 2" in the text.

Morton's racial category	Morton's 1849 average cranial capacities (in^3)	Reconstruction of the 1849 cranial capacity table (in^3)
Modern Caucasians	85	85
Mongolians	82	84
Malays	85	85
Americans	79	83
Africans	83	83

Column 1 is based on Morton's summary tabulations in his 1849 *Catalogue*. Column 2 was calculated as follows:

1. Modern Caucasians and Malay values are based on Morton's own averages, which Gould takes no issue with. The Caucasian value reflects Gould's original tabulation from the *Catalogue* summary table, which was correct.
2. The African value excludes the Australians and the Hottentots following Morton. Since Morton appears to have collected the Australians and Hottentots well after 1839, these were unlikely to have been part of his 1839 sample. So despite their relatively small cranial capacities, their exclusion here cannot explain the abnormally small African measurements in *Crania Americana*.
3. The American value of 83 reflects lifting Gould's arbitrary third criterion (restriction to the *Crania Americana* sub-sample) and including all Native American groups with at least three members. These data can be found in Lewis et al.'s Dataset S3, Table 9.
4. Gould notes (1996, 99) that the Mongolian value should be corrected by adding back a very large Chinese skull (#1336) with cranial capacity 98 in^3 that Morton seems to have omitted from his 1849 summary tabulation. Gould also argued that three Eskimo crania described in *Crania America* should have been added to this sample. Morton borrowed these skulls from a friend, and didn't remeasure them, which is presumably why he didn't include them in his 1849 tabulations. Thus, I believe we should include cranium #1336, but not try to include measurements for skulls Morton no longer possessed or measured. I have tried to consistently apply the following rules for inclusion: (a) data should be reported in the 1849 *Catalogue*, (b) sub-families, when reported, should have more than three members, and (c) all families and sub-families should be weighted equally.