4 Craniometry and Indigenous Repatriation

Cressida Fforde, Alexandra Roginski, Alan Goodman, Paul Turnbull, and Hilary Howes

Introduction

This chapter considers the use of morphometric analysis of the cranium ('craniometry' or 'craniometrics') for determining ancestry in repatriation cases. The use of craniometrics in repatriation appears to be quite widespread but, with some exceptions (e.g. Ousley, Billeck and Hollinger 2005; Wright 2008), its use is little chronicled and appears to be uncritiqued. Most applications of craniometry in repatriation processes are carried out in confidential consultancies (Wright 2008 p 111), and there is thus a lack of availability of cases for peer review. A potentially concerning aspect of craniometry's application in repatriation is the way in which its proponents may articulate its ability to determine population affinity as definitive, without raising awareness to clients of the criticisms that have been levelled against it by other members of the scientific fraternity. Another pressing issue is that the use of the technique, and in particular the way that results are communicated to a lay audience, may serve to reify, rather than deconstruct, the existence of 'race', leading to perpetuation of harmful narratives. In this chapter, we review scientific assessments on the utility of craniometrics in general to determine population affinity and then discuss a range of questions that arise from the way in which this technique appears to have been scaffolded into repatriation research, particularly from forensic science.

Our main aims are to provide repatriation practitioners with further information with which to assess craniometrics, its influence and the information it produces. We also wish to raise greater awareness about its use in repatriation so that this type of application can be included within discussions about the use of the technique more generally. It is interesting to explore its use in repatriation contexts for a range of other reasons. For example, the need for more discussion about the redundant notions of, and assumptions surrounding, race and racial traits that it may reinforce; broader questions surrounding understanding of the relationship (or not) between biological and social identity; the discrepancy it reveals about how sources of information should be weighted and 'believed'; and the implications of all these issues for Indigenous control, decision-making and free, prior and informed consent in the repatriation of Ancestral Remains.

Craniometry plays a vexed role in the analysis of human remains. Proponents of cranial studies continue to study and develop programmes and techniques to

DOI: 10.4324/9781003144953-4

hopefully discriminate amongst groups to aid in ancestry determination both to extend understanding of human history and migration, for example, but also in the very applied realm of forensic science and repatriation. This occurs despite persistent criticism from experts in the same disciplines that the accuracy of these methodologies is limited at best, should not be relied upon as evidence and at worse can uphold deeply problematic notions about race. In a Letter to the Editor published in the *Journal of Forensic Sciences*, Bethard and DiGangi (2020) succinctly identified the dangers associated with morphoscopic (identified) cranial traits, leading them to call for 'an immediate moratorium on the use of morphoscopic cranial traits in the estimation of ancestry given the lack of comprehensive inquiry into why the traits exist and the fact that *their use serves to bolster the debunked biological race concept*' (2020 p 1791, our emphasis). In 2021, they extended this critique 'to craniometric and dental morphological analysis in ancestry estimation' (DiGangi and Bethard 2021 p 424).

The arrival of Ancient DNA analysis (aDNA) added further critique to existing criticism of the accuracy of craniometrics to determine ancestry, and, in the USA, in ways that had a profound impact on repatriation in high profile cases (see Roginski and Fforde, this volume). Thus, in the case of The Ancient One (Kennewick Man) and the individual found in the Spirit Cave in Nevada (Hockett and Palus 2018; Rasmussen et al. 2015 pp 2–4) morphometric analysis had been used as evidence to block their return to Indigenous groups. Partly based on craniometrics, scientists argued that these ancient remains could not be claimed as Native American in accordance with the requirements of the USA's national repatriation legislation (NAGPRA). However, aDNA research later affirmed Native American ancestry, resulting in the repatriation of these individuals to claimant groups (Huber 2020). The use of craniometrics as a repatriation 'adjudicator' raises interesting questions about scientific practice in the return of Ancestral Remains. For example, it raises questions about who controls decision-making and what information they draw upon to determine 'truth', particularly because this occurs at a time when the fields of bioarchaeology and bioanthropology are shifting towards partnership and consultation with Indigenous peoples as recognised best practice (Boutin et al. 2017; Kakaliouras 2017, 2019 p 193).

Craniometrics in Repatriation

Central to the repatriation process is some level of certainty as to the identity of the deceased, necessary, of course, to ensure the right Ancestral Remains are returned to the right people. However, who determines what is 'right', what types of evidences are used in making such a determination and what weight these various types of evidence are given is not straightforward.

Identity determination can have bearing on repatriation processes in two distinct and different ways (see also Fforde, Keeler and Aranui 2020). The first is to determine whether the requested human remains are connected to those making the claim and thus are 'eligible' for repatriation. The second is to assist in locating more specifically who and/or where to return them to. In the first, the Ancestral

Remains are not under Indigenous control and institutions can, of course, be more or less in favour of repatriation – and thus more or less likely to prioritise information that casts doubt on Indigenous ancestry. The second occurs when Ancestral Remains are wholly or partially under Indigenous control and the question is not whether they should be returned or not, but how to assist in increasing levels of provenance where documentary evidence is non-specific or in the rare cases in which it is contradictory.

Craniometrics has been used in both scenarios. In the first, identity determination is generally at the population level – in reality, this is frequently whether the individual is believed to belong to the same race/ethnicity as the claimant (and see below for a discussion on terminology and the problem of craniometry acting to affirm the existence of racial categories). In the second, the use of craniometrics has been much more fine-grained. For example, in Australia, it has sought to link the deceased to particular Indigenous nations. Because of the very limited public information available about the use of craniometrics for the second purpose, this paper restricts its discussion to the use of craniometrics to determine ancestry at the macro-level of race/ethnicity and, concomitantly, its use in determining whether human remains are to be returned or not.

The level of historical documentation associated with museum-curated human remains varies. It is sometimes very detailed (enabling identification of locality of removal, burial site and sometimes even named individuals), often less specific (allowing regional provenance only – which can vary from a mountain range, river or bay, to a geographical area defined by colonial boundaries (for example a local council area, or a state, province or territory)) and sometimes very general (being restricted to a racial category or country of origin identification only). In a very few cases, the historical record may be confusing, contradictory and unclear, for example containing discrepancies that offer two different ancestry determinations for the deceased. Whether craniometric analysis is used as part of processes to document ancestry in repatriation processes depends upon the practices of each museum or repatriation agency (Indigenous or otherwise).

While it might be thought that craniometrics would be deployed only in cases where the historical record was absent or unclear, this is not so. Instead, proponents point to the potential inaccuracy of the historical record as one reason why morphological assessment is required as a standard practice in repatriation (Ousley, Billeck and Hollinger 2005 p 16). Rarely is any justification given for determination of such inaccuracy beyond conflicting craniometric results (although see Wright 2008). In our observation, craniometric analysis has only been used by holding institutions to either validate identity recorded in the archives or to refute it and thus remove particular human remains from the frame that are ‘claimable’. To our knowledge, morphological analysis has never been used to identify for Indigenous repatriation purposes those remains which have no numbering, labelling or documentation whatsoever (and whose identity is thus undocumented in the historical record) or those who have been ascribed a non-Indigenous identity in a museum catalogue.

As Kakaliouras observed in 2008, technical discussions within the disciplines of physical anthropology, bioarchaeology and forensic anthropology veered into the realm of sociology with the rise of repatriation practices (Kakaliouras 2008 p 45). In the USA, physical anthropologists have applied craniometry in response to the requirements of NAGPRA legislation (Ousley, Billeck and Hollinger 2005), under which claimant groups (until 2010, with the exception of the Smithsonian which is exempt from this requirement) had to show cultural affiliation to the remains. In Europe, some institutions have used craniometrics in their repatriation decision-making. Thus, at least one institution in Germany has refused to return a significant proportion of the human remains documented with Australian provenance in its catalogues, because a craniometry report cast doubt on this affiliation. In another German case, clear historical documentation identified the remains of a named and known individual. However, craniometric analysis produced a different result which, despite the detailed documentation, was given sufficient weight to cast doubt on the individual's Aboriginal identity. In the end, this individual was repatriated because the skeleton exhibited a fracture which was described in a letter by Hermann Klaatsch (1863–1916), a leading German anthropologist, who examined his remains in the early twentieth century. In Australia, where the museum sector since the early 1990s has been generally proactive in respecting the rights of Indigenous groups to take back control over ancestral remains, collections managers have used craniometrics with the hope it will assist in establishing the provenance of collections where documentation has been lost or sites of collection were not adequately recorded in the first place (Roginski 2015 p 9).

Terminology

Biological anthropologist John Relethford (2017 paragraph 3) notes that, following a shift in anthropology during the mid-twentieth century away from fixed typologies garnered through observation, some researchers turned their focus towards 'clusters of local populations in local aggregates', with attention paid to evolutionary forces therein. Relethford argues that craniometric variation, along with genetic differences, shows a clear geographic structure, but contends that 'the boundaries in global variation are not abrupt and do not fit a strict view of the race concept'. The lines are arbitrary and 'at best a crude first-order approximation' (Relethford 2009 p 16).

The word 'race' is no longer in common use in the field of bioarchaeology, with researchers generally using 'population' to refer to differences between groups located at a geographic distance (Kakaliouras 2003 p 2). In this field, as well as in forensic science, 'ethnicity' and 'ancestry' are often used. However, in practice, these terms have often been employed as if they were synonymous with race.

In forensic science, for example in missing persons cases, the reality is that police in the USA (and elsewhere) are concerned to determine basic information on age, race and sex of the deceased. The continuation of the term and concept of 'race' in forensic anthropology (and see Weiss and Springer 2020) despite its having been replaced in other areas of physical anthropology has been the subject of some scholarship and critique (see for example Sauer 1992 and response by Goodman 1997 and see also Bethard and DiGangi 2020; DiGangi and Bethard 2021). The enduring interchangeable nature of these terms was evident in a key 1990s edited volume by the *Bulletin of the National Association for the Practice of Anthropology*, in which Claire Gordon referred to forensic methods for referencing 'race/ethnicity' (Gordon 1993).[1]

In the context of forensic anthropology, identifying a deceased individual's race is one of several factors, along with sex and age, that scientists are expected to discern from unidentified human remains. During the mid-to-late twentieth century, the categories outlined in various law enforcement systems in the USA still included a tripartite division – such as 'Caucasoid', 'Mongoloid' and 'Negroid', or 'White', 'Negro' and 'American Indian' – with forensic anthropologists duly expected to make a determination amongst these racialised options (see above and, for example, Hinkes 1993; Snow et al. 1979 p 48). As Stanley Rhine explained in 1993, even if a practitioner might think that these categories were an oversimplification, 'the forensic anthropologist who delivers a philosophical lecture to the sheriff on the non-existence of human races is unlikely to be consulted again' (Rhine 1993 p 55). In contrast to this position, Smay and Armelagos (2000 p 20) argued that, for some anthropologists, the seeming utility of such categories proves their validity. Bethard and DiGangi (2020; see also DiGangi and Bethard 2021) have once more identified the problem inherent in forensic anthropology's continuing adherence both to the concept of race and its measurable evidence in cranial features, and the unquestioned justification of using morphometrics to identify racial affiliation for use in missing person cases. This need for members of the anthropological discipline to continually 'push back' against the acceptance of 'race' in forensic science reminds us of Ashley Montagu's barely contained exasperation at having to respond in the 1960s within the pages of *Current Anthropology* to continuing debate about the biological reality of 'race' and, in particular, the connection between brain size and intelligence:

> does it have to be re-proven every year that brain size within the normal range of variation characteristic of the human species at the *sapiens* level and characteristic of every human population has nothing whatever to do with intelligence?
>
> (Dobzhansky, Montagu and Coon 1963 p 362)

Measuring Skulls – A Brief History

In Chapter 3, Paul Turnbull provides a comprehensive history of skull measuring in the long nineteenth century for the purposes of understanding human origins and diversity. Here a brief history is provided in order to contextualise the current chapter, with a focus on the twentieth century.

The study of human skulls with the intent of understanding difference and perceived hierarchies has a long history. Early observations that different peoples could be distinguished by the shape of their heads date back to the Ancient Greeks at least. In the modern era, scientific interest is earliest seen in the work of Carl Linnaeus (1707–1778), Johann Friedrich Blumenbach (1752–1840), John Hunter (1728–1793) and Petrus Camper (1722–1789). While Blumenbach and Hunter observed differences in skull shape, Camper introduced measurements of skull features – notably facial slope (or Camper's Angle) – as a means of differentiating between races and, in doing so, established a technical means of geometrically classifying humankind into Blumenbach's schema of five geographically located varietal types – Caucasoid, Mongoloid, Ethiopian, American and Malay. Thereafter, new craniometric techniques were devised by key practitioners during the nineteenth century, such as physician-scientist Samuel George Morton (1799–1851). Morton, for example, measured the volume and shape of cranial cavities to produce a ranking of races by intelligence. He also considered what he took to be racial typicalities in the structural morphology of skulls to confirm the polygenetic origins of humanity.

The obsession with the size and shape of skulls carried forward into the twentieth century, notably in the research of Ales Hrdlička (1869–1943), founder of physical anthropology at the Smithsonian Institution and the first editor of the *American Journal of Physical Anthropology*. Hrdlička received training in how anthropometric analysis could be applied in a medico-legal context and became one of the key drivers for the development of forensic anthropology in the USA (Armelagos and Van Gerven 2003 pp 54–56; Hefner et al. 2016 pp 4–5; Ubelaker 2006 p 201).

During the early twentieth century, Franz Boas posed a challenge to claims that morphological traits, especially the cranial index (width/breadth), were stable over time. In his famous study of immigrants Boas' data showed that the children of Jewish and other European migrants had different head shapes to those of their parents (Boas 1912). Instead of supporting the assumption of stable 'types', Boas proposed that skull shape had a degree of 'plasticity' that responded to different environmental conditions (Boas 1912 p 557). Boas is widely cited in anthropology as one of the first scientific critiques of the explanatory value of the concept of race. However, his study had little influence on the significant uptake of craniometrics in the applied sciences, notably forensic anthropology. Re-evaluations of Boas' work in the early 2000s were undertaken by Sparks and Jantz (2003) and Gravlee, Bernard and Leonard (2003a). Interestingly, the authors came to two different conclusions – the former that Boas 'got it wrong' and the latter that he 'got it right'. Gravlee, Bernard and Leonard (2003b) suggest why two groups of scientists looking at the same data could come to such different conclusions. The divergent conclusions are an indication that adherence or rejection of a fundamental pillar upon which rests the utility of craniometrics for determining population affinity is to do with much more than simply 'crunching the numbers'.

Through the twentieth century, skeletal biologists and others developed a wide range of measurements and observable traits to aid in the determination of race (e.g. Gill and Rhine 1990). In the US context, race was typically categorised as European/White, African/Black and Native American. Traits included observations on the shape of postcranial remains, such as the pelvis and femur, measurements of these same postcranial elements, and similar observations of dental discrete traits and tooth sizes. As in prior centuries, the greatest attention was placed on skulls, including discrete traits such as the size of the brow ridges, the shape of the nose and the shape of the mastoid process.

Since the 1960s, the analysis of single skull dimensions, ratios such as the cranial index, and the subjective analysis of shapes have largely been supplanted by methods that consider multiple measures of size and shape. These then rely on multivariate statistics to determine the best fit or fits of a skull to a group, based on statistical distance measurements.

The first and best-tested method was developed by Eugene Giles and Elliot (1962) and is most often referred to as some variant of the Giles and Elliot method. Giles and Elliot recognised that a newly developed statistical technique called multivariate linear discriminant function could sort skulls into 'sex' and 'races'. In essence, this technique, now aided by computers, finds the single equation that best divides or discriminates two samples into their appropriate groups. The computer programme is designed to determine the best combination of measurements and constants. The programme's aim is to properly discriminate or choose between two previously chosen groups. A drawback is that it cannot distinguish three or more groups at the same time. Thus, for three 'races' three equations are needed: one to distinguish the first from the second, the second to distinguish the second from the third and the third to distinguish the first from the third. The answer to the multiple regression equations are continuous variables. However, in developing the equation, an optimal break point is ascertained. Values above the break point are assigned to one group and values below the point to the second group. The break point is simply that point at which the highest percentage of 'cases' is correctly placed.

In the USA, in the early 1960s, Giles and Elliot, then graduate students working under W.W. Howells at Harvard, decided to use discriminant function analysis to try and distinguish the racial categories of White, Negro and American Indian from a set of metric measurements. Large samples of skulls of individuals whose sex and race had been documented were needed. To this end, researchers turned to various collections where such information was recorded. They first developed equations that discriminated males and females (Giles and Elliot 1963). Then they developed a set of equations that best separated male Whites from American Indians and Whites from Negro males and similarly female Whites from American Indians and Whites from Negro females. In this way, a tripartite race separation was made for each sex.[2] They found that their equations generally identified the correct 'race' about 85–90% of the time. This high percentage, however, is not unexpected as it was based on the equation that worked best on the submitted samples.

Giles and Elliot (1962), recognising that circularity – that the method would work well simply because they had chosen the equation that worked best – smartly left over some skulls (75 of each 'race/sex') and re-ran the equation on these skulls. The equations did less well on the remaining samples. For example, the percent correct of Indian males dropped from 95% to 77%.

The re-test problem amplifies when samples are tested from other locations. It is unusual for a forensic method to undergo a re-study or re-test. However, this method is quite central to the field, a sort of gold standard, and Goodman (1997) was able to locate four re-tests of Native American crania in the literature. He finds that the average correct identification of 'race' drops to 33% or exactly random. He postulates that the reasons why include greater variation within 'races' than between 'races' and that the cultural definitions of 'race' change over time and space. The lack of replicability of these studies is a good example of just how unstable and unreliable 'race' categories are.

To try to overcome some of the limitations of the Giles and Elliot method, in the 1990s new methods were developed, of which Fordisc is probably the most widely used (see below). While it is more sophisticated than the Giles and Elliot method, Fordisc and similar programmes such as CRANID suffer from the same fundamental issues, including great variation within populations or 'races' and changes in cultural definitions.

Modern Craniometry: Methods and Critique

The question of whether cranial studies can identify morphological distinctions between groups with high rates of success underpins assessment of the potential value of these methods in the provenancing of human remains for repatriation purposes. The different methodologies that have emerged for cranial identification in the forensic, archaeological and repatriation settings include those based on anatomical examination, those applying multivariate measurements and those that focus on prominent characteristics identified as being more statistically significant in particular populations. These methodologies have all attracted criticism.

Aside from determining population affiliation, craniometry and other methods of analysis also seek to establish sex, a parameter for which there is no clear binary. As Zindel (2015 p vi) observes, 'Human variation is endless, and even in areas of the skeleton for which it has been established that there is a significant degree of sexual dimorphism, there will be individuals who do not fit neatly into a binary conception of sexual divergence'. Physical anthropologists also assess whether sex can be judged from other bones, including the clavicle, ulna and femur (for example, see Kearns 2015) and especially the pelvis.

As noted above, the digitised application of multivariate analysis has dominated craniometrics since the 1960s. Multivariate analysis assigns skulls to a sex and then a race/population, according to the best fit derived from a series of measurements between 'landmarks' on the skull. After obtaining a series of measurements from each skull, these measures are entered into a computer programme and the fit of the measurement is compared to those of reference samples from different populations to generate a statistical assessment of similarity (see Dudzik and Kolatorowicz 2016 for an explanation of these different sets of landmarks, and the histories and contexts of their creation).

The US software package Fordisc, developed by Stephen Ousley and Richard Jantz and first released in 1992, incorporates the US Forensic Anthropology Data Bank, as well as data collected by US archaeologist William Howells (Elliot 2008 p 3). Meanwhile, CRANID, developed by Richard Wright of the University of

Sydney, uses a different statistical method, as well as the Howells sample. CRANID is claimed to be more appropriate for Australia and Europe because it contains data from more Aboriginal and European human remains in its reference sample (Dudzik and Kolatorowicz 2016 p 55; Kallenberger and Pilbrow 2012 p 460). Unlike the analysis of Giles and Elliot (1962) by Goodman (1997), there have been very few published re-tests of these newer packages. Yet, both of these packages have attracted criticism (see below) by practitioners who have tested their efficacy using remains of known provenance (see, for example, Elliot and Collard 2009; Kallenberger and Pilbrow 2012; Konigsberg, Algee-Hewitt and Steadman 2009; Smay and Armelagos 2000; Williams, Belcher and Armelagos 2005; and for some discussion see also Hubbe et al. 2007; Keita et al. 2007). Both are used in repatriation cases.

The field burgeons with other software and digital tools for morphometric assessment, although Fordisc and CRANID have the strongest hold. These other packages include Osteoware, the Smithsonian Institution's data-entry software package that can capture a range of qualitative and quantitative data, including macromorphoscopics (Osteoware 2011). The rASUDAS application utilises 21 dental morphological traits developed as part of the Arizona State University Dental Anthropology System to estimate ancestry (Scott et al. 2018). The Portugese programme AncesTrees, which also uses the Howells dataset, is based on the algorithmic method of 'random forest' modelling and was inspired by Fordisc and CRANID (Navega et al. 2015). This machine-learning method builds on the algorithmic system developed by Hefner and colleagues (Hefner, Spradley and Anderson 2014; Navega et al. 2015 p 1146). The programme 3D-ID promises to glean ancestry and sex from the shapes of skulls, thus improving on 'linear distances, ratios, and angular measurements that generally lose valuable shape information' (3D-ID website 2019; Thomas 2017 p 27). A recent thesis (Skalic 2018 p 20) compared AncesTrees and Fordisc 3.1, using nine discriminant functions, and found that both programmes were incorrect for determining race in more than 50% of cases.

Williams, Belcher and Armelagos (2005 p 341) note that Fordisc 2.0 uses 'social race as an approximation of biological affinity' and argue that it constructs 'population and racial traits [as] constant over time and space' and that its proponents use it uncritically, even when it reveals results contrary to expectations. Williams, Belcher and Armelagos (2005) tested Fordisc by measuring and entering the details of 42 Nubian crania. While Nubians were not represented in the Fordisc 2.0 reference sample, the researchers nevertheless expected that they would group with ancient Egyptians, who are genetically more like each other than any other group. Instead, amongst the sample, only eight were grouped with Egypt, six as Zalavar (Hungary), four as Easter Island, three as Lake Alexandrina (Australian Aboriginal) and three others as Norse (Williams, Belcher and Armelagos 2005 p 342). These authors contend that Fordisc does not account for the heterogeneity of human morphology within any given population. 'Even a presumably homogenous population such as the Meroitic Nubians shows extensive variation that preclude its classification as a geographic group' they write (2005 p 345). They also argue that Fordisc applies an American idea of race that does not account for how race is socially constructed in international settings and does not consider how folk taxonomies of race will vary by nation and culture (Williams, Belcher and Armelagos 2005 p 344).

In a similar project, Naar, Hilgenberg and Armelagos (2006) tested the efficacy of Fordisc for assessing the ethnicity of Egyptian crania, for which a reference sample does exist. Applying crania from within Howells' population sample, they found that the programme only correctly identified 55 of 111 Egyptian crania when researchers used the statistical parameters outlined by Freid et al. (2005). Naar et al. therefore concluded that 'Cranial variation, even in one of the core samples, is too great to be reliably classified as belonging to that sample' (2006 p 136).

Marina Elliot's (2005) thesis assessed the efficacy of Fordisc using the dataset created by Howells. Without entering source population or sex as an informed prior (see below), the programme achieved a best result of 36.5%. Elliot concluded that, 'As it stands, Fordisc requires the population, the time period, the sex and as many measurements as possible for a set of remains before it can be expected to return a reasonable estimation of ancestry' and that, 'Given this situation, the only conclusion that can be drawn is that if Fordisc is used at all, it should only be under extremely restricted circumstances or to provide limited confirmation of information gathered through other means' (2005 p 77).

With co-author Mark Collard, Elliot also concluded that Fordisc should only be used in limited circumstances, when an unidentified specimen is 'more or less complete' and where such a specimen belongs to one of the populations represented by the programme (Elliot and Collard 2009 p 849). Elliot and Collard also critiqued the seminal dataset created by Howells, noting that he himself admitted that some of the remains may have had sex misattributed and that *he only chose remains that he deemed typical for a particular group*, thus subjectively exaggerating the characteristics that he asserted defined that population (2009 p 851). This stated selectivity fundamentally undermines the utility of what is perceived to be a valid 'reference' sample. Finally, Aronsen and Ellis (2009) attempted to use Fordisc 3.0 to provenance remains in Yale's Peabody Museum of Natural History, but the programme mis-categorised two of the four crania in their control sample (assigning Japanese identity to one Australian and one Indian cranium).

Even supporters of Fordisc acknowledge that it does not work well outside of the USA because of fewer reference populations (Manthey et al. 2018 p 263.e1). In the European context, forensic scientists Pierre Guyomarc'h and Bruzek (2011) applied Fordisc to French and Thai collections of remains in order to test the efficacy of the programme for determining sex. They found that the rate of accuracy varied between 52.2% and 77.8%, 'depending on the options and groups selected' and advised that 'primary sexual diagnosis using cranial measurements must only be made when a reference sample is available' (Guyomarc'h and Bruzek 2011 p 180.e5). They rejected the use of the programme for European forensic work (Guyomarc'h and Bruzek 2011 p 180.e5). Ramsthaler, Kreutz and Verhoff (2007) also tested the efficacy of Fordisc for sex determination using a German forensic collection of 98 crania and found a higher probability that Fordisc could diagnose sex appropriately, with a result of 94% for males and 86% for females (Ramsthaler, Kreutz and Verhoff 2007 p 477). Ramsthaler et al. nevertheless concluded that Fordisc should not be used without an update to include European reference samples (2007 p 477).

Fordisc has also proven to have limitations in the South African context, with Rebecca King (2015 p v) – in a comparison of Fordisc with geometric morphometric programme 3D–ID – finding respective accuracy rates of 75.6% and 63.1%. King attributed this result to low numbers of South African individuals within the comparative reference sample, but with no way to test this explanation it can only remain a hypothesis. Physical anthropologist Curtis Wienker (undated p 4) has tested the efficacy of Fordisc using Indian, white Cuban and Chinese Cuban skulls, finding that 'the results of these Fordisc analyses are a mixture of a little accuracy, lots of relative inaccuracy and considerable inconsistency'.

Even bioanthropologists who find Fordisc a useful tool – such as Pamela Geller and Stojanowski (2017), who applied it for cross-checking the identities of seven individuals thought to be of Seminole ancestry in the Morton collection – point out its limitations for establishing the distinctions between specific population groups. 'Our analysis shows that it is difficult to differentiate between African and Native American affiliations as the comparative samples become more modern in age' (Geller and Stojanowski 2017 p 280), a factor that Geller and Stojanowski attribute to 'gene flow' between these groups in Florida. Jodi Blumenfeld already noted the complication of 'admixture' in the US forensic setting in 2000 (Blumenfeld 2000 p 20).

Fordisc's developers argue that critics do not use the programme appropriately or that it isn't advised for use with remains without a corresponding reference population (Elliot 2008 p 2). Elliot observes that 'it seems that each time a criticism of the program is raised, Fordisc's developers add a new caveat to its use' (2008 p 74). In response to the Nubian study of Williams et al. (2005), Ousley, Jantz and Freid (2009 p 73) argued that the scholars did not use a sufficient number of variables to achieve a satisfactory result and that the overlap between population groups is well acknowledged in physical anthropology. Williams and Armelagos (in Hubbe et al. 2007), in defending the 2006 paper by Naar et al., pointed out that the 12 variables used by researchers align with one of the examples in the Fordisc 2.0 manual and that 'a greater number of variables will not universally contribute to a better classification in discriminant-function analysis' (Hubbe et al. 2007 p 286).

Part of what is interesting and telling about these failed analyses is the way in which their researchers attribute such failure to factors such as incorrect measurements, admixture, gene flow or a poor reference sample. Such attributions indicate a much greater readiness to explain failure through external factors, rather than a willingness to question whether the central method (and supporting tenets) are instead fundamentally defective – even when others have done so. It is not unreasonable to compare these 'explanations of failure' to those sought by craniologists of the late nineteenth and early twentieth centuries when their own statistical analyses of skull measurements started to show, completely unexpectedly, greater variation within the 'races' than between them. The initial response to this unexpected result was to explain it through miscegenation, distance from racial 'type' and small sample size. Craniologists sought to address these issues by taking even more measurements and collecting even more skulls. It was only later that scientists began to question the validity of race as a biological construct, and even then it was only with the evident horrors of Nazi behaviour in the Second World War that any kind of consensus on the matter was reached via the UNESCO (1952) Statement on Race. As this chapter shows, 'race' is an extremely tenacious concept. Thomas Kuhn's (1962) work on scientific revolutions allows some insight into why this may be the case.

CRANID

CRANID has also been used in a number of repatriation cases. CRANID has attracted less assessment than Fordisc, but has received mixed responses. Lauren Kallenberger and Pilbrow (2012 p 459) studied 23 skulls of known origin and found that CRANID assigned 39% of skulls 'to geographically closest matching reference samples and 48% to major geographic regions'. In a test of the very principle that race can be gleaned from skeletal characteristics, Sierp and Henneberg (2015) analysed 20 unprovenanced skulls using nine recognised methods for determining race, ranging from the seminal triracial division of Giles and Elliot (see above) to the multivariate analysis of CRANID. For none of the 20 skulls did all of the findings correlate in one known 'race'. Sierp and Henneberg noted that 'Fourteen of the skulls were identified as belonging to all three of the racial classes' (2015 p 8). Of CRANID, they said that the programme will not work if even one of the 29 anthropometric points cannot be measured, and 'if the most likely sample population has a high probability and if the sample populations that immediately follow are consistently from contiguous geographical areas' (2015 p 29). Other researchers have investigated the subjective way in which measurements are taken. Richard Slater and Martin Smith, who see CRANID as a valuable tool, have observed high rates of variability amongst the results of 11 practitioners tasked with obtaining measurements for one skull (from a London Quaker burial ground), with the variation in measurements meaning that the skull was grouped by CRANID with ten different populations, ranging from Bushman, to Punjabi, to Danish Neolithic and to Italian (Slater and Smith 2012 p 137).

Appropriate Scaffolding into Repatriation? Informed Priors, Evidence and Estimation

As well as the overall issue of whether database packages such as Fordisc and CRANID can reliably determine population affinity and the associated danger of supporting a racial logic, there is the question of whether a method devised largely for the purpose of identification of individuals in forensics is automatically suitable for use in repatriation research. One way to interrogate this question is the use of what is known as 'informed priors'.

As Konigsberg, Algee-Hewitt and Steadman (2009) showed in their description of a particular forensic case study, the applied statistical analysis undertaken by Fordisc can return different results depending upon whether an 'informed prior' is introduced or not, and also what type of 'prior' is used. It is worth explaining this case study in more detail as it is informative about how craniometrics is used in forensics, and the resulting additional potential dangers of scaffolding this technique into repatriation. The work of Konigsberg, Algee-Hewitt and Steadman (2009) indicates that the 'job' that craniometrics is asked to do in forensics is subtly – but significantly – different to that which it is required to do in repatriation; there is an important distinction to be made in assessment of its value for *estimatory* purposes (forensics) and in its value for *evidentiary* purposes (repatriation).

In the case study considered by Konigsberg et al. (2009), a forensic anthropologist was brought a skull known to have been found in Iowa. When compared with all datasets (the global dataset) in Fordisc (i.e. no prior information was introduced), the programme returned that the skull was that of an Easter Islander. The forensic anthropologist knew that this identification was unlikely to be the case, so instead compared it to what he/she deduced were populations in the database that were similar to what the census data said were the demographics of Iowa. i.e. the forensic anthropologist introduced 'prior' information. Note that census data is based on self-identification to various 'categories' (there are questions of course as to whether the 'group' people self-identify as is the same 'group' that their 'biology' is *presumed* to link them to). When this was done, Fordisc returned that the skull belonged to the majority group in Iowa – 'American White'. However, amply illustrating the problem with Fordisc and the power of 'informed priors', when this 'prior' information was changed to a different population, the ancestry determination also shifted: 'it would have been classified as a Pacific Islander had his remains been found on Hawaii and as an "American Black" had his remains been found in Gary, Indiana' (Konigsberg et al. 2009 p 83).

Because of this type of lack of certainty, Konigsberg et al. (2009) agree with others who are much more critical of Fordisc that the evidentiary power of Fordisc is minimal and its use in forensic anthropology should be for estimation only.

Within forensic science, this issue of informed priors may not be critical as the main aim is to provide the police with a way to identify *possibilities* within databases containing thousands of missing person cases. Each case in the reduced pool then comes with many other ways of ultimately identifying the deceased. However, using the technique as evidence for determining ancestry in repatriation cases is to take the dependence on craniometrics to a much higher level. There appears no standard procedure for use of informed priors in repatriation cases, nor discussion about the veracity or otherwise of the source of that prior information. However, its importance is clearly critical. For example, if a skull documented as 'Australian' in the historical record is measured and compared to a global dataset in Fordisc and a result returned that determines considerable distance from the metrics of Australian human remains in the reference package, is it then standard procedure to introduce 'Australia' as an informed prior?

Head Shape, Ancestry Determination and Identity in Repatriation

Ultimately, the utility of Fordisc and CRANID to determine ancestry is based on the premise that head shape can determine ancestry and hence geographical origin, notwithstanding the now widely accepted view that the greatest majority of variation is intra-group. Over and above the issues of accuracy discussed above, this premise leads to a distinct issue in how results are interpreted and communicated. For example, rather than stating (a) that the target remain is statistically similar in shape to, say, skulls from Vietnam, a report may state (b) that the target remain is *from* Vietnam (or *is* Vietnamese). Statements (a) and (b) are not the same thing. The first represents what the statistical analysis has shown. The second is a conclusion based on the premise that human groups/geography can be identified on the basis of skull shape. Rarely, if ever, is such a distinction articulated clearly. By identifying this issue in the way results are articulated, it is easy to see again how craniometrics can be articulated (and understood) in a racial frame.

The shortcomings of, and implications surrounding, the use of terminology are apparent in the case of The Ancient One (Kennewick Man) which opened this paper. In this case, the failings of the discipline have been publicly aired in relation to the category of the 'Paleoamerican' population, a classification created by physical anthropologists in North America who argued that some ancient remains were so unlike modern-day Native Americans in terms of skull shape as to support a hypothesis that another ancient group, characterised instead by 'Caucasoid' features, had resided in North America (Jantz and Owsley 2001 p 153; Kakaliouras 2019). Physical anthropologists including Douglas Owsley of the Smithsonian Institution classified the 8,500+-year-old remains of The Ancient One (Kennewick Man) in Washington state and those found in Spirit Cave in Nevada, as Paleoamerican, rather than Paleoindian (Kakaliouras 2019). As noted above, in the case of The Ancient One, this evidence convinced a judge that the requirement of cultural affiliation could not be fulfilled in line with NAGPRA legislation (Burke et al. 2008). The eventual refinement of methods for extracting DNA from ancient remains (Matisoo-Smith and Horsburgh 2012 Chapter 1) led to research that disproved the claim (Rasmussen et al. 2015). Importantly for the purposes of this section of the paper, these findings showed that genetic lineage does not correlate with skeletal typologies, bringing further discredit to the notion of ancestry determination through analysis of head shape.

Nonetheless, it is equally important to move away from the notion that biology (whether suggested by aDNA or morphology) and ancestry are one and the same. As Masur points out, 'the relationship between language family, cultural practices, material culture, skeletal morphology, and geographic region is a complex and politically charged issue today, much as it was a century ago' (Masur 2013 p 4). Kakaliouras (2003 pp iv, 4) critiques a common assumption that biology and language developed along similar trajectories. She writes: 'it is important to consider that biological distance is a direct measure of human morphology and genetic relationships, and only an indirect source for piecing together patterns of culture, ethnicity or language in the present or distant past, especially at the regional level'

(2003 p 9). A view which considers morphology and group identification to be one and the same finds it difficult to accommodate social identity in all its variety and fluidity. Instead it may continue to be most 'comfortable' with considering that there is 'racial type' of skull shape. In this way both craniometry and DNA analysis and, in particular, the way in which their results are communicated can replicate highly problematic narratives that affirm a racial logic. Goodman (2007) writes about three types of ancestry – genetic, cultural and geographic – and notes that for a wide range of reasons the three may or may not line up.

Mixed Ancestry

A related issue that requires some unpacking is the concept of 'admixture', a term used to describe the influence on an individual's skeletal morphology of having ancestors from more than one (usually two) groups (usually 'racial' categories). The term, its function and the assumptions and perceptions about identity (and authenticity) in which it is embedded form important context for decisions as to whether craniometrics is an appropriate tool to deploy in repatriation practice and assessment of related information that may result. Fordisc, CRANID and other statistical packages are most comfortable with assigning population affinity in a world in which (usually imagined as existing at some time-period in the past) populations show unique morphology and are bounded and discrete. Reliant on the fundamental tenets that head shape is unique to a population shows greater similarly within these populations to those externally, and is largely unchanging through time, the skulls of anonymous individuals, it is reasoned, can be assigned population affinity based on similarly of skull shape to those present in a reference sample. From a technical point of view, on the one hand, individuals who bring ancestry from different populations confound this reasoning, but on the other 'admixture' also appears to be a catch-all explanation for results that are not clear cut – for example, showing similarity to two or more populations.

A search of publications using Fordisc and CRANID for determining ancestry shows how commonly such reasoning is used to explain unclear results (e.g. Hughes et al. 2019). However, perhaps surprisingly, there appear to be very few studies which focus on trying to understand (and thus predict) the influence of 'mixed ancestry' on skeletal morphology, and whether identification of admixture is justified or even possible for ascribing identity using craniometric methods (although see Go et al. 2019; Olson 1999; Urbanová et al. 2014 for some attempts at the latter). It may be much more likely that unclear results are not due to 'admixture' as claimed, but simply natural variation; but this result would fundamentally question the utility of database packages such as Fordisc and CRANID. Furthermore, and just as importantly, use of terms such as 'admixture' can act to affirm (certainly it does not deny) underlying assumptions about 'race' and 'racial purity' which continue to have significant negative impact and have underpinned a suite of highly socially, culturally and individually damaging regimes experienced by Indigenous peoples and others, the legacy of which continues today.

One of the many questions that arise out of this shortcoming for repatriation purposes is whether, and how much, discordant results between statistical packages and the historical record may relate to the mixed ancestry of the deceased and, if so, what the consequences are and how such a possibility can be elevated in current decision-making by holding institutions. For, of course, whether or not the deceased does have, for example, Aboriginal and European ancestry says nothing about how they themselves identified or were identified by their community and has little relevance to the question of whether or not their remains should be returned. It also exposes a distinct flaw in the type of reasoning which sees supposed racial origin as a defining factor in repatriation, rather than the connection of the deceased to those who claim them, illustrating the complexity inherent in the connection between 'identity' and 'rights to decide'.

Ultimately, the statistical possibilities for error render craniometry inappropriate for the specific nature of repatriation, in which the consequences of decisions about provenance can have serious repercussions. As detailed above, in the forensic anthropology context, Konigsberg et al. (2009 p 78) argue that the statistical success rate of morphological methods, though imperfect, may still qualify these techniques for use as an additional tool for *estimating* the potential identity of an unknown person. But cranial analysis in the repatriation setting appears to commonly take on the quality of privileged *evidence* rather than merely assisting in estimation, as demonstrated by the legal impact of the Kennewick Man case. Even if a high rate of success was possible (as seen, for example, in Klales and Kenyhercz 2015) and may seem statistically compelling in a research paper, a 5% or 10% chance of error in assigning an individual to a group is significant if it means that remains are returned to the wrong people, or withheld erroneously from claimants.

Hierarchies of Evidence

Despite these documented shortcomings, craniometry is still perceived by its adherents as an objective scientific technique. In conjunction with the dominance that is frequently conferred on science as arbiter of 'truth', this means that the results of craniometrical analyses tend to be considered more reliable and are thus accorded more weight, than the results of historical research. As mentioned above, holding institutions adhering to these hierarchies of evidence have been known to refuse to return human remains if the provenance recorded in archival documents does not match the affiliation suggested by a craniometry report or to raise concerns that might cause obstacles in the repatriation claim process. This section explores this type of situation in more depth.

Ayau et al. (this volume) detail how such a situation occurred and was managed, in the case of Native Hawaiian ancestral remains housed at the Duckworth Laboratory at Cambridge University until they were repatriation in 2019. In an earlier case, the Natural History Museum (NHM) in London withheld eight human remains which were documented as Native Hawaiian in its own catalogues, because of evidence provided by Fordisc. The NHM had returned nearly 150 Native Hawaiian ancestral remains as a result of the Museum's own provenancing processes which included craniometric and historical research. This case is one in which a museum's desire to retain remains was likely not an incentive for 'over-favouring' craniometric data. Instead, it may relate to intrinsic notions concerning what type of data should be given more weight in decision-making. Even in cases where a holding institution is favourably disposed towards repatriation, craniometric data that conflicts with historical information can thus disrupt the repatriation process by sowing doubt in the minds of curators, consultant provenance researchers, government representatives and claimants. Both forensic anthropologists and historians are called on as 'expert witnesses' in provenance research, much as they are in legal cases (Evans 2002; Haack 2008; Petrovič 2009). In our observation, however, the prevailing assumption amongst policymakers with regard to provenance research is that doubts introduced by craniometric analyses should be given priority over even detailed and consistent historical records. In none of these cases is explanation given about why the historical documentation is considered faulty, apart from the presence of contradictory craniometric results. Of course, the tension between the sciences and the humanities – particularly in any area concerning human cultures, society and history – is not new.

At first glance, this tendency to prioritise scientific evidence over historical evidence challenges much of what has been written in recent years about the 'post-truth society' and declining public trust in the scientific enterprise (e.g. Farrell 2020; Iyengar and Massey 2019; Lockie 2017; Lynch 2020; Spector 2020; Weingart 2002). It also complicates research identifying ideological and practical barriers to the uptake of scientific consensus in policymaking (e.g. Baekgaard et al. 2017; Haas 2004; Young et al. 2002). It does, however, fit very neatly into what is generally termed 'scientism', described by Hyslop-Margison and Naseem (2007 p 117) as 'a belief that scientific knowledge is the foundation of all knowledge and that, consequently, scientific argument should always be weighed far more heavily than other forms of argument'. The 'other forms of argument' most often considered by critics of scientism are those derived from philosophy, especially moral and ethical philosophy (e.g. Burnett 2018; de Ridder, Peels and van Woudenberg 2018; Williams and Robinson 2015), but several critics have also acknowledged scientism's devaluation of knowledge derived from history and the humanities more broadly (Kitcher 2012; Kleinberg 2016; Scruton 2013).

This devaluation hinges on misperceptions of both scientific and historical research. Whereas scientific research is commonly perceived as objective, rigorous, and resulting in 'logically indisputable proofs about the natural world', historical research is seen as subjective, lacking in rigour, and able at best to deal effectively with particularities, but not to produce large-scale conclusions (Oreskes 2004 p 369; see also Kitcher 2012; Kleinberg 2016). It is seldom appreciated that different scientific enterprises are 'remarkably various in their methods' (Kitcher 2012) and that 'a uniquely effective method or procedure – the supposed "scientific method"'– does not in fact exist, at least not in the sense of a single, unchanging method used by all scientists and only by scientists (Haack 2012 pp 87–89; see also Haack 2003). Rather than producing incontrovertible and timeless truths, science is better understood as resulting in 'a robust consensus based on a process of inquiry that allows for continued scrutiny, re-examination, and revision' (Oreskes 2004 pp 369–370; see also Duschl 2020 pp 188–189; Gould 1981 pp 21–23). Indeed, a fundamental aspect of the principles of the scientific method is to constantly review and reassess techniques in light of new findings. Adherence to models which have consistently accrued strong criticism points to more complex and subjective reasons why they continue to be utilised.

Popular depictions of the scientific endeavour rarely reflect these types of nuances (Bowdoin Van Riper 2003). Much has been written about the 'CSI effect', the likely impact of popular television programmes such as *CSI: Crime Scene Investigation* that depict forensic evidence as infallible. Scholars unanimously agree that such programmes grossly overstate the accuracy and certainty of forensic science, though they are divided as to the impacts of such exaggeration on jurors' decisions (e.g. Cavender and Deutsch 2007; Hughes and Magers 2007; Machado and Granja 2020 pp 45–56; Ogilvie 2011; Podlas 2006; Salmon 2008; Smith 2016). In addition, scholars from a range of disciplinary backgrounds have emphasised the relationship, actual and perceived, between measurement and knowledge, 'the special status that numbers enjoy' as 'the ultimate test of objectivity' and the resulting 'overuse of measurement' in virtually every domain of human activity (Gould 1981 p 26; Henshaw 2006; see also Hyslop-Margison and Naseem 2007; Williamson and Piattoeva 2019). Taken together, popular depictions of scientific infallibility and prevalent assumptions of the superiority of qualitative data may help explain the social weight given to findings produced by craniometric analyses (or other types of scientific work) over outcomes produced by historical research.

Approaches to Assessing Accuracy in the Archive

Having assessed the information provided by craniometrics, it is important to likewise assess that which is provided by relevant historical sources. A useful way to approach this question is to ask what factors might lead to inaccuracy. On the one hand, those donating ancestral remains to institutions, or collecting on their behalf, knew the importance of ensuring crania and other bodily structures were those of people believed to be of 'pure' racial ancestry. Importantly, they also understood that the scientists who received the remains would be able to tell the difference. Donors regarded themselves as contributors to scientific knowledge of the human condition, and often sought by their donations to nurture intellectual and often wider social relations with prominent scientists. Sending falsely attributed remains would have jeopardised these relationships, and the donor's perceived reliability and credibility. Hence donors would seem to have little incentive to falsify racial ancestry or lineage. In 1877, so concerned was the Rev. Joseph King that he may have sent European remains in error that he wrote to George Rolleston, Professor of Anatomy at Oxford, 'I hope it is the skull of a bona fide aboriginal that I have sent you. I discovered the other day that some white settlers were buried about twenty years ago near to the spot were [sic] the skull was obtained' (Rolleston Papers, Ashmolean Library, Oxford).

Similarly, professional collectors risked their livelihoods if the items they sent were not correctly classified – whether Ancestral Remains, or other items provided to museums. Professional and amateur scientists who engaged in collecting, as well as institutionally contracted or employed collectors, often documented the circumstances in which they were obtained and affirmed their identity. For example, William Lowthian Green, a businessman and, later, senior public official in Hawai`i, sent numerous Native Hawaiian ancestral remains to UK collections in the mid-nineteenth century, taking care to describe where they were obtained, and cautioning that in two instances he was ‘Not quite certain where found but have no doubt they are native skulls’ (Green 1856). Likewise, the professional naturalist James Wilcox, who collected for the National Museum of Victoria in the mid-1860s, made sure to confirm that skulls he sent were ‘characteristic and authentic’ (Wilcox 1865).

On the other hand, while there is little to suggest that scientists or museum curators were concerned *en masse* that they were receiving large quantities of falsely or inaccurately provenanced remains, there is some evidence that craniologists could be wary of particular sources and the asserted identity of Ancestral Remains from particular places (see Knapman, Howes and Winkelmann this volume). However, concluding because of such concerns that they were indeed receiving inaccurately identified remains – and assessing how much this should affect repatriation claims is not straightforward. This is for a number of reasons, not least because of the belief that skulls of certain people should look a certain way. If assessment of authenticity was based on skull shape, then, as we have explained above, such assessments hold little weight. In Chapter 6, Howes, Knapman and Winkelman explore three case studies in which the authenticity of human remains was debated by craniologists of the day. Interestingly, the views of the craniologists were not uniform, and their motivation for claiming inauthenticity is varied. For repatriation practitioners today, it would be remiss to reject descriptions in the archive, particularly if these are detailed, unless there was very clear evidence (from historical and other sources) to support an assessment that such information was incorrect.

Where there is a greater risk to the accuracy of provenance is when museum processes, possibly occurring at various times over many decades, compromise information. Generally, although not in every case, nineteenth-century museum registration processes are thorough and leave a comprehensive paper trail. Cases of mistaken identity which have occurred post-registration can usually be identified if all surviving museum documentation is consulted. In well-documented collections, cases of misidentification are rare, which suggests that they are also likely to be rare in collections where documentation has been lost. Provenance research may also be hindered by collections having been moved between institutions, or in other ways becoming separated from associated archives. However, in these cases the result is usually that human remains become anonymous, rather than wrongly identified.

Overall, there seems little reason to doubt the veracity of the great majority of historical documentation, or the potential of good historical research to identify errors and discrepancies in the attribution of provenance. Finally, we would do well to reflect that databases such as Fordisc and CRANID are comprised of measurements of skulls with provenances confirmed by associated documentation. Hence should historical documentation be regarded as inherently suspect, what are the implications for the use of these measurements in these software packages?

Conclusion

In 1861, the British ethnologist John Crawfurd lamented that 'a vast amount of anatomical skill and labour has been exercised upon the skull, in the hope – always defeated – of making it a distinctive test of race' (Crawfurd 1861 p 374). Some 70 years later, Arthur Thomson, Oxford anatomist and anthropologist, observed in his lectures on physical anthropology that the,

> diversity of skull form displayed in the human species is such that hitherto all attempts to classify have failed. It is easy to make broad generalizations, but … these … are in this case somewhat misleading. On the other hand, if you attempt to go into refinements, the complexity becomes so great that practically the system becomes unworkable.
>
> (Thomson 1931 p 205)

It appears that little has changed to revise these assessments. The lack of opportunity for peer-review of craniometric methods in repatriation decreases confidence still further.

In a 2020 letter to the editor of the *Journal of Forensic Science*, forensic anthropologists Jonathan Bethard and Elizabeth DiGangi called for a 'moratorium on the use of morphoscopic cranial traits in the estimation of ancestry given the lack of comprehensive inquiry into why the traits exist and the fact that their use serves to bolster the debunked biological race concept' (Bethard and DiGangi 2020 p 1791). In 2021 they have responded decisively to the resulting commentary (Stull et al. 2021) and have used Critical Race Theory to continue their critique of the use of morphoscopic traits and expand this exploration to craniometric and dental morphological analysis (DiGangi and Bethard 2021). While discussing the veracity of such techniques for forensic utility, their concomitant argument identifies forensic anthropology as a discipline 'cloaked in whiteness' and centres around the societal implications of a method that upholds popular notions of race (DiGangi and Bethard 2021 pp 6–7). While they do not consider the use of such techniques in repatriation, their findings of the 'parent discipline' are equally applicable. Thus:

> in addition to the discussed scientific problems, the main point is not whether or not we are consciously or purposefully perpetuating the biological race concept, or whether ancestry estimation 'works,' or whether researchers have created more sophisticated ways to demonstrate that it works –the point here

> is that by providing an ancestry estimate grounded in traits of the skull, we are reinforcing law enforcement and the public's belief in the concept of biological race. And this contribution in and of itself is harmful.
>
> (DiGangi and Bethard 2021 p 6)

Their intervention in 2021, almost 70 years after the UNESCO Statement on Race (1952), speaks to the tenacity of the race concept and the role of science in its maintenance. A recent book by Weiss and Springer (2020) that opposes repatriation uses the term race (noting that its use has 'become problematic for some' (2020 p 6)) and believes the term 'appropriate in the forensic context' (2020 p 6), also highlights its endurance at high levels of academia. The Weiss and Springer volume (2020) and the very need for DiGangi and Bethard's (2021) recent analysis is evidence that forensic anthropology 'has not done the work to truly reject the biological race concept and we are so delinquent with this task that we do not even recognize race science when it is staring us in the face' (DiGangi and Bethard 2021 p 431).

When deciding whether to scaffold craniometrics into repatriation, it is of the utmost importance that this broader context is understood. Research for repatriation practice is a new area of applied research that can potentially employ many techniques used in the forensic sciences. However, using those that have questionable value and risk perpetuating notions of biological race – a prime intellectual driver of the collecting of Indigenous remains in the first place – would be a retrograde step and needs very careful thought. As discussed in Fforde, Keeler and Aranui (2020), research for repatriation practice is fundamentally about determining identity. This raises two important questions: first, what techniques can be used to determine identity; and secondly, who determines the evidence 'bar' and makes the ultimate decision about identity and whether to repatriate or not. And as discussed above, we encounter the paradoxical situation that craniometry has been used to dispute identity and refuse repatriation claims, despite clear contrary historical documentation, but has yet to be used to activate repatriation for remains where no provenance information at all survives.

Repatriation is fundamentally about resetting a colonial relationship between Indigenous nations and the museums and associated disciplines that took, exhibited and studied their remains through the devastating lens of racial science. Craniometrics is a technique with its roots strongly embedded in this colonial history, which is vulnerable to employing concepts such as race as if they were biological realities rather than social constructs. As discussed above, the analytical vocabulary employed in craniometrics (e.g. 'admixture') can too closely echo the language of later nineteenth-century anthropology. That craniometrics continues to be used to determine ancestry in areas of anthropology that engage most directly with popular and historical notions of race is informative. The technique can be used to provide simple (yet heavily critiqued) answers to complex issues of human diversity and identity. But while seemingly attractive to solve difficult matters (whether in the forensic sciences or repatriation), these answers have proven to be problematic and can easily re-substantiate mistruths and discourses of difference that constrain cross-cultural relations and have pernicious effects on the socio-economic conditions, health and well-being of those most affected by colonialism and racial science. The use of craniometrics in repatriation runs the risk of validating the same kinds of typological reasoning about human variation that characterised the race science of old. And if it helps repatriation claims by validating information in historical documentation, it is important to recognise that in doing so it presents the danger of reviving, normalising and perpetuating that most dangerous myth of race.

Notes

1 'Ancestry' features as a common term for biological connection in a more recent handbook by Pilloud and Hefner (2016a). Rather than attempting to assign people to one of a handful of races, 'biological distance' (or 'biodistance') studies are often concerned with relationships between populations on a small scale (Hefner et al. 2016 p 11). Pilloud and Hefner (2016b p xxiv) argue that biodistance analysis perceives the skeleton as 'a dynamic tissue affected by multiple factors throughout life. Consideration must be given to the influence of stress, age, activity levels, diet, and disease, and when possible controlled'. Biodistance scholars such as Hefner et al. (2016 p 13) argue that most practitioners are familiar with the historical harm caused by racial classification. Nevertheless, citing Ousley et al. (2009), they contend that 'there is a demonstrable correlation between social race, skeletal morphology, and ancestry. This concordance is likely due to genetic drift, assortative mating, institutional racism, and geographic patterning among US populations' (Hefner et al. 2016 p 13). Within a museum setting, researchers may apply physical anthropology to attempt to discern the population to which particular remains are most closely related (Ousley et al. 2005 pp 6–7).

2 Eugene Giles told one of us (AHG) that he had the good fortune of using a new computer at Harvard. He punched out cards (the data) and the programme was submitted to a computer operator who would then submit the cards to the computer and with luck he would get his results the day after.

References

3D-ID website. 2019. Retrieved 14 February 2019 from https://www.3d-id.org/.

Armelagos, George J. and Van Gerven, Dennis P. 2003. A century of skeletal biology and paleopathology: Contrasts, contradictions, and conflicts. *American Anthropologist*, *105*(1), pp. 53–64.

Aronsen, G.P. and Ellis, S.C. 2009. Identification of unprovenienced human remains using Fordisc 3.0: Biological race revisited. *American Journal of Physical Anthropology*, *138*(S48), pp. 79–80.

Baekgaard, Martin, Christensen, Julian, Dahlmann, Casper Mondrup, Mathiasen, Asbjørn and Petersen, Niels Bjørn Grund. 2017. The role of evidence in politics: Motivated reasoning and persuasion among politicians. *British Journal of Political Science*, *49*, pp. 1117–1140.

Bethard, J. and DiGangi, E. 2020. Letter to the Editor – moving beyond a lost cause: Forensic anthropology and ancestry estimates in the United States. *Journal of Forensic Sciences*, *65*(5), pp. 1791–1792.

Blumenfeld, Jodi. 2000. Racial identification in the skull and teeth. *Totem*, *8*(1), pp. 20–33.

Boas, Franz. 1912. Changes in the bodily form of descendants of immigrants. *American Anthropologist*, *14*(3), pp. 530–562.

Boutin, Alexis T., Long, Madison, Dinarte, Rudy A. and Thompson, Erica R. 2017. Building a better bioarchaeology through community collaboration. *Bioarchaeology International*, *1*(3–4), pp. 191–204.

Bowdoin Van Riper, A. 2003. What the public thinks it knows about science: Popular culture and its role in shaping the public's perception of science and scientists. *EMBO Reports*, *4*(12), pp. 1104–1107.

Burke, Heather, Smith, Claire, Lippert, Dorothy, Watkins, Joe and Zimmerman, Larry (eds). 2008. *Kennewick Man: Perspectives on the Ancient One*. Walnut Creek: Left Coast.

Burnett, Thomas. 2018. What is scientism? American Association for the Advancement of Science. Retrieved 14 January 2021 from https://www.aaas.org/programs/dialogue-science-ethics-and-religion/what-scientism.

Cavender, Gray and Deutsch, Sarah K. 2007. CSI and moral authority: The police and science. *Crime Media Culture*, *3*(1), pp. 67–81.

Crawfurd, J. 1861. On the classification of the races of man. *Transactions of the Ethnological Society of London*, *1*, pp. 354–378.

de Ridder, Jeroen, Peels, Rik and van Woudenberg, René (eds). 2018. *Scientism: Prospects and Problems*. Oxford: Oxford University Press.

DiGangi, E.A. and Bethard, J.D. 2021. Uncloaking a lost cause: Decolonizing ancestry estimation in the United States. *American Journal of Physical Anthropology*, *175*, pp. 422–436.

Dobzhansky, Theodosius, Montagu, Ashley and Coon, C.S. 1963. Two views of Coon's origin of races with comments by Coon and replies. *Current Anthropology*, *4*(4), pp. 360–367.

Dudzik, B. and Kolatorowicz, A. 2016. Craniometric data analysis and estimation of biodistance. In Pilloud, Marin and Hefner, Joseph (eds), *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*. Amsterdam: Elsevier Science, pp. 35–60.

Duschl, Richard A. 2020. Practical reasoning and decision making in science: Struggles for truth. *Educational Psychologist*, *55*(3), pp. 187–192.

Elliot, Marina. 2008. *Fordisc and the determination of ancestry from craniometric data*. PhD diss., Simon Fraser University.

Elliot, Marina and Collard, Mark. 2009. Fordisc and the determination of ancestry from cranial measurements. *Biology Letters*, *5*(6), pp. 849–852.

Evans, Richard J. 2002. History, memory, and the law: The historian as expert witness. *History and Theory*, *41*(3), pp. 326–345.

Farrell, Katharine N. 2020. Untrol: Post-truth and the new normal of post-normal science. *Social Epistemology*, *34*(4), pp. 330–345.

Fforde, C., Keeler, H. and Aranui, A. 2020. Research for repatriation practice. In Fforde, C., McKeown, C.T. and Keeler, H. (eds), *The Routledge Companion to Indigenous Repatriation: Return, Reconcile, Renew*. Abingdon: Routledge, pp. 541–563.

Fforde, Cressida, McKeown, C. Timothy, Keeler, Honor, Ormond-Parker, Lyndon, Tapsell, Paul, Turnbull, Paul, Hemming, Steve, Rigney, Daryle, Pickering, Michael, Aranui, Amber, Morris, Wes and Knapman, Gareth. 2021. Identity in applied repatriation research and practice. In Meloche, C.H., Spake, L. and Nichols, K.L. (eds), *Working with and for Ancestors: Collaboration in the Care and Study of Ancestral Remains*. Abingdon, Oxon: Routledge, pp. 255–267.

Freid, Donna, Spradley, M.K., Jantz, Richard L. and Ousley, Steve D. 2005. The truth is out there: How NOT to use FORDISC. *American Journal of Physical Anthropology*, *126*(S40), p. 103.

Geller, Pamela L. and Stojanowski, Christopher M. 2017. The vanishing Black Indian: Revisiting craniometry and historic collections. *American Journal of Physical Anthropology*, *162*(2), pp. 267–284.

Giles, Eugene and Elliot, Orville. 1962. Race identification from cranial measurements. *Journal of Forensic Sciences*, *7*, pp. 147–157.

Giles, Eugene and Elliot, Orville. 1963. Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology*, *21*, pp. 53–68.

Gill, G.W. and Rhine, S. (eds). 1990. *Skeletal Attribution of Race*. *Anthropological Papers No. 4*. Albuquerque: University of New Mexico, Maxwell Museum of Anthropology.

Go, M.C., Jones, A.R., Algee-Hewitt, B.F., Dudzik, B. and Hughes, C.E. 2019. Classification trends among contemporary Filipino crania using Fordisc 3.1. *Forensic Anthropology*, *2*(4), pp. 293–303.

Goodman, Alan. 2007. Toward genetics in an era of anthropology. *American Ethnologist*, *34*(2), pp. 227–229.

Goodman, Alan H. 1997. Bred in the bone? *The Sciences*, *37*(2), pp. 20–25.

Gordon, Claire. 1993. Why classify race? *NAPA Bulletin*, *13*(1), pp. 1–6.

Gould, Stephen Jay. 1981. *The Mismeasure of Man*. New York and London: W.W. Norton.

Gravlee, C.C., Bernard, H.R. and Leonard, W.R. 2003a. Heredity, environment, and cranial form: A reanalysis of Boas's immigrant data. *American Anthropologist*, *105*(1), pp. 125–138.

Gravlee, C.C., Bernard, H.R. and Leonard, W.R. 2003b. Boas's changes in bodily form: The immigrant study, cranial plasticity, and Boas's physical anthropology. *American Anthropologist*, *105*(2), pp. 326–332.

Green, W.L. 1856. *'List and Account of Skulls of Sandwich Islanders'* pasted within *Catalogue of human crania &c in the collection of Joseph Barnard Davis, Volumes I–IV Compiled c. 1857–c.1878*, NB. Items in the collection are numbered 1-1818 inclusive. Vol. I Nos. 1–481; Vol. II Nos. 482–886; Vol. III Nos. 887–1336; Vol. IV Nos. 1337–1818; Original manuscript volumes in the Library of the Royal College of Surgeons of England; current call nos: 42.c.35–42.c.38 Microfilmed Nov. 1999

Guyomarc'h, Pierre and Bruzek, Jaroslav. 2011. Accuracy and reliability in sex determination from skulls: A comparison of Fordisc® 3.0 and the discriminant function analysis. *Forensic Science International*, *208*, pp. 180.e1–180.e6.

Haack, Susan. 2003. *Defending Science – Within Reason: Between Scientism and Cynicism*. Amherst: Prometheus.

Haack, Susan. 2008. Of truth, in science and in law. *Brooklyn Law Review*, *73*(3), pp. 985–1008.

Haack, Susan. 2012. Six signs of scientism. *Logos & Episteme*, *3*(1), pp. 75–95.

Haas, Peter. 2004. When does power listen to truth? A constructivist approach to the policy process. *Journal of European Public Policy*, *11*(4), pp. 569–592.

Hefner, J.T., Pilloud, M.A., Buikstra, E. and Vogelsberg, C.C.M. 2016. A brief history of biological distance analysis. In Pilloud, Marin and Hefner, Joseph (eds), *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*. Amsterdam: Elsevier Science, pp. 3–22.

Hefner, J.T., Spradley, K. and Anderson, B. 2014. Ancestry assessment using random forest modeling. *Journal of Forensic Sciences*, *59*(3), pp. 583–589.

Henshaw, John M. 2006. *Does Measurement Measure Up? How Numbers Reveal and Conceal the Truth*. Baltimore: Johns Hopkins University Press.

Hinkes, Madeleine J. 1993. Race, ethnicity, and forensic anthropology: Realities of racial determination in a forensic setting. *NAPA Bulletin*, *13*(1), pp. 48–53.

Hockett, Bryan and Palus, Emily. 2018. A brief history and perspective on Spirit Cave, Nevada. *Paleoamerica*, *4*(1), pp. 1–7.

Hubbe, Mark, Neves, Walter A., Williams, Frank L'Engle and Armelagos, George J. 2007. On the misclassification of human crania: Are there any implications for assumptions about human variation? *Current Anthropology*, *48*(2), pp. 285–288.

Huber, A. 2020. Genomic testing of ancient DNA: The case of the ancient one (also known as Kennewick Man). In Fforde, C., McKeown, C.T. and Keeler, H. (eds), *The Routledge Companion to Indigenous Repatriation: Return, Reconcile, Renew*. Abingdon, Oxon: Routledge, pp. 625–636.

Hughes, C.E., Dudzik, B., Algee-Hewitt, B.F., Jones, A. and Anderson, B.E. 2019. Understanding (mis)classification trends of Latin Americans in Fordisc 3.1: Incorporating cranial morphology, microgeographic origin, and admixture proportions for interpretation. *Journal of Forensic Sciences*, *64*(2), pp. 353–366.

Hughes, Thomas and Magers, Megan. 2007. The perceived impact of crime scene investigation shows on the administration of justice. *Journal of Criminal Justice and Popular Culture*, *14*(3), pp. 259–276.

Hyslop-Margison, Emery J. and Naseem, M. Ayaz. 2007. *Scientism and Education: Empirical Research as Neo-Liberal Ideology*. Dordrecht: Springer.

Iyengar, Shanto and Massey, Douglas L. 2019. Scientific communication in a post-truth society. *Proceedings of the National Academy of Sciences (PNAS)*, *116*(16), pp. 7656–7661.

Jantz, Richard and Owsley, Douglas. 2001. Variation among early North American crania. *American Journal of Physical Anthropology*, *114*, pp. 146–155.

Kakaliouras, Ann. 2003. *Biological distances and the ethnolinguistic classification of Late Woodland (AD 800–1650) Native Americans on the coast of North Carolina*. PhD diss., University of North Carolina.

Kakaliouras, Ann. 2008. Leaving few bones unturned: Recent work on repatriation by osteologists. *American Anthropologist*, *110*(1), pp. 44–52.

Kakaliouras, Ann. 2017. NAGPRA and repatriation in the twenty-first century: Shifting the discourse from benefits to responsibilities. *Bioarchaeology International*, *1*(3–4), pp. 183–190.

Kakaliouras, Ann. 2019. The repatriation of the Palaeoamericans: Kennewick Man/The Ancient One and the end of a non-Indian ancient North America. *British Journal for the History of Science Themes*, *4*, pp. 79–98.

Kallenberger, Lauren and Pilbrow, Varsha. 2012. Using CRANID to test the population affinity of known crania. *Journal of Anatomy*, *221*(5), pp. 459–464.

Kearns, Aisling. 2015. *An evaluation of metric method for sex estimation using the clavicle, humerus, radius and ulna*. PhD diss., Boston University.

Keita, S.O.Y., Williams, Frank L'Engle, Belcher, Robert L. and Armelagos, George J. 2007. On Meroitic Nubian crania, Fordisc 2.0, and human biological history. *Current Anthropology*, *48*(3), pp. 425–427.

King, Rebecca Elizabeth. 2015. *Estimating ancestry in South Africa: A comparison of geometric morphometrics and traditional craniometrics*. PhD diss., Boston University.

Kitcher, Philip. 2012. The trouble with scientism: Why history and the humanities are also a form of knowledge. *New Republic*, 4 May. Retrieved 14 January 2021 from https://newrepublic.com/article/103086/scientism-humanities-knowledge-theory-everything-arts-science.

Klales, Alexandra R. and Kenyhercz, Michael W. 2015. Morphological assessment of ancestry using cranial macromorphoscopics. *Journal of Forensic Sciences*, *60*(1), pp. 13–20.

Kleinberg, Ethan. 2016. Just the facts: The fantasy of a historical science. *History of the Present*, *6*(1), pp. 87–103.

Konigsberg, Lyle W., Algee-Hewitt, Bridget F.B. and Steadman, Dawnie Wolfe. 2009. Estimation and evidence in forensic anthropology: Sex and race. *American Journal of Physical Anthropology*, *139*(1), pp. 77–90.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lockie, Stewart. 2017. Post-truth politics and the social sciences. *Environmental Sociology*, *3*(1), pp. 1–5.

Lynch, Michael. 2020. We have never been anti-science: Reflections on science wars and post-truth. *Engaging Science, Technology, and Society*, *6*, pp. 49–57.

Machado, Helen and Granja, Rafaela. 2020. *Forensic Genetics in the Governance of Crime*. Singapore: Palgrave Macmillan.

Manthey, Laura, Jantz, Richard L., Vitale, Albarita and Cattaneo, Cristina. 2018. Population specific data improves Fordisc's performance in Italians. *Forensic Science International*, *292*, pp. 263.e1–263.e7.

Masur, Laura Elizabeth. 2013. *Virginia Indians, NAGPRA, and cultural affiliation: Revisiting identities and boundaries in the Chesapeake*. PhD diss., The College of William and Mary.

Matisoo-Smith, Elizabeth and Horsburgh, K. Ann. 2012. *DNA for Archaeologists*. Ebook. New York: Routledge.

Naar, Nichole A., Hilgenberg, Dana and Armelagos, George J. 2006. Fordisc 2.0 the ultimate test: What is the truth? *American Journal of Physical Anthropology*, *129*(S42), p. 136.

Navega, D., Coelho, C., Vicente, R., Ferreira, M.T., Wasterlain, S. and Cunha, E. 2015. AncesTrees: Ancestry estimation with randomized decision trees. *International Journal of Legal Medicine*, *129*(5), pp. 1145–1153.

Ogilvie, Emily M. 2011. *Fake forensics, real effects? Testing the cultivating power of crime drama*. MA diss., University of North Carolina at Chapel Hill.

Olson, D.D. 1999. *Admixture and racial classification: The use of discriminant analysis in classifying individuals of mixed ancestry*. Masters diss., University of Montana.

Oreskes, Naomi. 2004. Science and public policy: What's proof got to do with it? *Environmental Science and Policy*, *7*, pp. 369–383.

Osteoware, Smithsonian. 2011. Retrieved 14 February 2019 from http://osteoware.si.edu/.

Ousley, Stephen D., Billeck, William T. and Hollinger, R. Eric. 2005. Federal repatriation legislation and the role of physical anthropology in repatriation. *American Journal of Physical Anthropology*, *128*(S41), pp. 2–32.

Ousley, Stephen D., Jantz, Richard and Freid, Donna. 2009. Understanding race and human variation: Why forensic anthropologists are good at identifying race. *American Journal of Physical Anthropology*, *139*(1), pp. 68–76.

Petrovič, Vladimir. 2009. *Historians as expert witnesses in the age of extremes*. PhD diss., Central European University.

Pilloud, Marin and Hefner, Joseph (eds). 2016a. *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*. Amsterdam: Elsevier.

Pilloud, Marin and Hefner, Joseph. 2016b. Preface. In Pilloud, Marin and Hefner, Joseph (eds), *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*. Amsterdam: Elsevier Science, pp. xxiii–xxvii.

Podlas, Kimberlianne. 2006. The CSE effect: Exposing the media myth. *Fordham Intellectual Property, Media & Entertainment Law Journal*, *16*(2), pp. 429–466.

Ramsthaler, F., Kreutz, K. and Verhoff, M.A. 2007. Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. *International Journal of Legal Medicine*, *121*(6), pp. 477–482.

Rasmussen, Morten, Sikora, Martin, Albrechtsen, Anders, Korneliussen, Thorfinn Sand, Moreno-Mayar, J. Victor, Poznik, G. David, Zollikofer, Christoph P.E. et al. 2015. The ancestry and affiliations of Kennewick Man. *Nature*, *523*, pp. 455–464.

Relethford, John. 2009. Race and global patterns of phenotypic variation. *American Journal of Physical Anthropology*, *139*(1), pp. 16–22.

Relethford, John H. 2017. Biological anthropology, population genetics, and race. In Zack, Naomi (ed.), *The Oxford Handbook of Philosophy and Race*, online edition. New York: Oxford University Press.

Rhine, Stanley. 1993. Skeletal criteria for racial attribution. *NAPA Bulletin*, *13*(1), pp. 54–67.

Roginski, Alexandra. 2015. *The Hanged Man and the Body Thief: Finding Lives in a Museum Mystery*. Clayton: Monash University Publishing.

Ryan, L. 1974. Report to the Australian Institute of Aboriginal Studies on Truganini. Unpublished manuscript. AIATSIS pMs 1436.

Salmon, Christine D. 2008. DNA is different: Implications of the public perception of DNA evidence on police interrogation methods. *Richmond Journal of Law and the Public Interest*, *11*(2), pp. 51–82.

Sauer, Norman J. 1992. Forensic anthropology and the concept of race: If races don't exist, why are forensic anthropologists so good at identifying them? *Social Science & Medicine*, *34*(2), pp. 107–111.

Scott, G. Richard, Pilloud, Marin, Navega, David, Coelho, João, Cunha, Eugénia and Irish, Joel. 2018. RASUDAS: A new web-based application for estimating ancestry from tooth morphology. *Forensic Anthropology*, *1*(1), pp. 18–31.

Scruton, Roger. 2013. Scientism in the arts and humanities. *New Atlantis*, *40*, pp. 33–46.

Sierp, Ingrid and Henneberg, Maciej. 2015. Can ancestry be consistently determined from the skeleton? *Anthropological Review*, *78*(1), pp. 21–31.

Skalic, Carmen Bryanne. 2018. *A critical approach to ancestry in forensic anthropology: an assessment of Fordisc 3.1 and AncesTrees*. PhD diss., University of Windsor.

Slater, Richard J. and Smith, Martin J. 2012. A study of interobserver variation in cranial measurements and the resulting consequences when analysed using CranID. In Mitchell, Piers D. and Buckberry, Jo (eds), *Proceedings of the Twelfth Annual Conference of the British Association for Biological Anthropology and Osteoarchaeology*. Oxford: Archaeopress, pp. 131–139.

Smay, Diana and Armelagos, George. 2000. Galileo wept: A critical assessment of the use of race in forensic anthropology. *Transforming Anthropology*, *9*(2), pp. 19–29.

Smith, Peter Andrey. 2016. When DNA implicates the innocent. *Scientific American*, *314*(6), pp. 11–12.

Snow, Clyde, Hartman, Steve, Giles, Eugene and Young, Fontaine. 1979. *Sex and Race Determination of Crania by Calipers and Computer: A Test of the Giles and Elliot Discriminant Functions in 52 Forensic Cases*. Virginia: National Technical Information Service.

Sparks, C.S. and Jantz, R.L. 2003. Changing times, changing faces: Franz Boas's immigrant study in modern perspective. *American Anthropologist*, *105*(2), pp. 333–337.

Spector, Bert. 2020. Post-truth claims and the wishing away of brute facts. *Leadership*, *16*(1), pp. 9–24.

Stull, K.E., Bartelink, E.J., Klales, A.R., Berg, G.E., Kenyhercz, M.W., L'Abbé, E.N., Go, M.C., McCormick, K. and Mariscal, C. 2021. Commentary on: Bethard JD, DiGangi EA. Letter to the Editor – moving beyond a lost cause: Forensic anthropology and ancestry estimates in the United States. *Journal of Forensic Science*, *65*(5), 1791–1792. *Journal of Forensic Sciences*, *66*(1), pp. 417–420.

Thomas, Alexandra N. 2017. *The estimation of ancestry and sex in unknown individuals through a comparison of methods*. MA diss., Florida Atlantic University.

Thomson, A. 1931. *Lectures on Physical Anthropology*, 2 vols. Oxford: Pitt Rivers Museum.

Ubelaker, Douglas H. 2006. Evolution of the relationship of forensic anthropology with physical anthropology and forensic pathology: A North American perspective. *Studies in Historical Anthropology*, *4*, pp. 199–205.

UNESCO. 1952. *The Race Concept: Results of an Inquiry*. Paris: United Nations Educational, Scientific and Cultural Organization (UNESCO).

Urbanová, P., Ross, A.H., Jurda, M. and Nogueira, M.I. 2014. Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample. *Journal of Legal Medicine (Tokyo)*, *16*(5), pp. 264–273.

Weingart, Peter. 2002. The moment of truth for science: The consequences of the 'knowledge society' for society and science. *European Molecular Biology Organization (EMBO) Reports*, *3*(8), pp. 704–706.

Weiss, E. and Springer, J.W. 2020. *Repatriation and Erasing the Past*. Gainesville: University Press of Florida.

Wienker, Curtis. Undated. Using biological data to challenge the reality of race. Unpublished paper. Retrieved 13 February 2019 from https://www.researchgate.net/publication/266453632_USING_BIOLOGICAL_DATA_TO_CHALLENGE_THE_REALITY_OF_RACE.

Wilcox, James F. 1865 Letter to XXX March 29, 1865. Museum Victoria Archives: correspondence 1854–1899, Alphabetic Box W.X.Y.Z.

Williams, Frank L'Engle, Belcher, Robert L. and Armelagos, George J. 2005. Forensic misclassification of ancient Nubian crania: Implications for assumptions about human variation.*Current Anthropology*, *46*(2), pp. 340–346.

Williams, Richard N. and Robinson, Daniel N. (eds). 2015. *Scientism: The New Orthodoxy*. London: Bloomsbury.

Williamson, Ben and Piattoeva, Nelli. 2019. Objectivity as standardization in data-scientific education policy, technology and governance. *Learning, Media and Technology*, *44*(1), pp. 64–76.

Wright, Richard. 2008. Detection of likely ancestry using CRANID. In Marc Oxenham (ed.), *Forensic Approaches to Death, Disaster and Abuse*. Bowen Hills: Australian Academic Press, pp. 111–122.

Young, Ken, Ashby, Deborah, Boaz, Annette and Grayson, Lesley. 2002. Social science and the evidence-based policy movement. *Social Policy and Society*, *1*(3), pp. 215–224.

Zindel, Greta. 2015. Masculinization of postmenopausal female crania: fact or fiction? PhD diss., Boston University.