VOLUME 57, No. 4, 482-715 **Summer 2017**

MANKIND

Quarterly

Mankind Quarterly is published by Ulster Institute for Social Research, London, UK
and printed in the USA.

 ISSN 0025-2344

CONTENTS

Mankind Quarterly is a refereed quarterly journal published by the Ulster Institute for Social Research, 28 Haycroft Gardens, London NW10 3BN, UK. Telephone +44 7 887 672 738. Fax +44 20 8181 4941. Email: info@ulsterinstitute.org. Web site: www.mankindquarterly.org. The price for a one-year subscription to the Print version is US$170.00 (plus US$23.00 postage to Non-US destinations), US$193.00 for the Electronic version or US$239.00 for a combined Print and Electronic subscription (plus US$23.00 postage to Non-US destinations). Electronic subscriptions are only available to Institutions with dedicated IP addresses, and there is an additional charge for access from multiple IP ranges beyond five. Private individuals may subscribe to the Print version at a privilege price of US$64.00 when mailed to their home addresses (plus US$23.00 for Non-US addresses). For further details, including payment methods, visit www.mankindquarterly.org/subscribe.

Cover picture: A market seller in Yakutsk, capital of the Siberian region of Yakutia, offers an Arctic hare for sale in addition to deep-frozen fish. Yakutsk is the coldest city (population > 100,000) on Earth, with an average winter temperature of -38ºC, and the only one of comparable size to be built on permafrost.
See the article by Vladimir Shibaev and Richard Lynn on page 680.

MANKIND QUARTERLY 2017 57:4 482-486

Editorial

Archaeogenetics: (Pre-)history Goes Molecular

One year ago, in the Summer 2016 issue of *Mankind Quarterly*, I wrote an editorial drawing the reader's attention to a "paradigm shift" in the social sciences, triggered by the discovery of genetic variants associated with educational attainment. Three large meta-analyses of genome-wide association studies (GWAS) had reported about one hundred regions in the genome in which common genetic variants (single-nucleotide polymorphisms, SNPs) are statistically associated with educational attainment. Each of these regions is inferred to contain a gene, with common variants in the gene affecting the chances of staying in school or earning an educational degree. It was obvious that suddenly, studies were becoming possible that had been unthinkable only a few years earlier. What happened to this "scientific revolution" during the last year? Did it produce any dramatic new insights, or was it all hype?

At the time, we already had first results from the United States showing that family advantage and social mobility are driven in part by "educational attainment genes" that children have inherited from their parents (Domingue et al., 2015). Could these results be confirmed in further studies? The most relevant new insights have come from a relatively small (N = 919) but thorough study in New Zealand (Belsky et al., 2016). It showed that a polygenic score for educational attainment was associated with numerous outcomes: (1) As expected, it predicted attainment of higher educational degrees. The effect was weak (r = .15), with a 19% increased risk of a university degree for a one standard deviation rise in polygenic score. (2) Children with higher polygenic scores showed accelerated cognitive development, had better self-control and better interpersonal skills. (3) Adult attainment was higher: more prestigious occupations, higher income, more accumulated assets, less welfare benefits, and higher credit scores. Only half of this was mediated by educational attainment. Thus the traits that are related to this polygenic score affect life outcomes not only by raising educational credentials. (4) Those with higher polygenic scores tended to come from families with higher socio-economic status, but higher polygenic scores were associated with upward social mobility for those from lower socio-economic backgrounds. (5) Adults with higher polygenic scores were more likely to emigrate or to have worked abroad. These results show compellingly that genetic variants predict many social outcomes, although with very low accuracy.

How do the genes affect these outcomes? Small effect sizes make studies of individual SNPs difficult, but the top SNP in the Rietveld et al. (2013) study turned out to be associated with verbal ability, with each allele contributing about 1.4 points to the verbal IQ (Gunnarsson et al., 2016). Therefore, this particular variant is an "IQ gene" that causes a verbal IQ difference of 2.8 points between the two homozygotes. More recent genome-wide association studies are finding "IQ genes" not by studying education-associated SNPs, but directly by genotyping subjects for whom IQ test scores have been obtained. The most recent study, based on results from 78,308 subjects, reports IQ-associated variants in 18 genomic regions, and an additional 30 genes that came out significant in gene-based analyses (Sniekers et al., 2017).

Genetic traits change their frequencies across generations through mutation, selection and random drift. Therefore the question arises of whether the frequencies of those genetic variants that contribute to individual differences in intelligence, personality or education have changed recently in human evolution. The question is, in particular, whether the long-term trend of rising cultural complexity is related not only to sporadic innovations and their cultural transmission, but also to changing frequencies of those genetic variants that make people capable of innovation and cultural transmission.

The first approach to this question was based on comparisons of allele frequencies for "educational attainment genes" and "IQ genes" in living human populations. Interested readers should consult the most recent publications by Davide Piffer (2016, 2017a,b,c) for the current status of this research. Using very different methods, Racimo, Berg & Pickerel (2017) have recently confirmed Piffer's conclusion that educational attainment genes have been subject to polygenic selection, at least in Europe and Asia. Their method even promises to show *when* selection for these genes has occurred.

Closer examination of the results shows that many of the SNPs, and most of the polygenic scores that are computed from the available SNPs, predict average intelligence and school achievement of modern human populations quite well, but others don't. This is expected because the currently available GWAS hits, though sometimes loosely described as "educational attainment genes" or "IQ genes", have a serious limitation: Most or all of them are *not* causal variants. They are statistically associated with the measured trait because they are located next to a hitherto unidentified causal variant on the chromosome. The two are in linkage disequilibrium because they are transmitted together through the generations until a meiotic crossing-over event separates them, which happens perhaps once every thousand generations or less depending on the distance between them on the chromosome and local recombination rates. Because of the rarity of crossing-over, since the time of the origin of the causal variant (or a later population bottleneck), one allele (variant) of the identified SNP is statistically associated with the low-IQ causal variant (presumed to be the ancestral one) while the other allele is associated with the high-IQ causal variant, (presumed to be of more recent origin). The observation of Piffer (2016) that only 55% of the "high-education alleles" in our collection of GWAS hits are derived and 45% are ancestral indicates that very few of the SNPs that we have so far are causal.

Therefore, the linkage phase between an identified SNP and the linked causal variant can be different in human populations. An SNP allele that is statistically associated with a "high-IQ" causal variant in Europeans may be associated with the alternative "low-IQ" causal variant in Asians or Africans. It is also likely that some causal alleles that are polymorphic (variable) in Europe are monomorphic (constant) elsewhere, and vice versa. Virtually all genome-wide association studies so far were done in people of European origin, and the resulting polygenic scores, for both medical and behavioral traits, are more predictive in Europeans than in other populations. For example, one American study observed that a polygenic score for educational attainment predicted 5.5% of the variance in educational outcomes in non-Hispanic Whites but only 1% in non-Hispanic Blacks (Ware et al., 2017). Therefore there is an urgent need for large-scale genome-wide association studies in non-European populations. Generally, Asians and Europeans are more suitable for microarray-based discovery studies, while Africans are more suitable for sequencing-based fine mapping because their genomes have lower levels of linkage disequilibrium.

Variable linkage phase is unlikely to confound studies that compare living populations with their ancestors. A first study of this kind has been performed with "educational attainment genes" in ancient European genomes dating back to the Bronze Age (Woodley of Menie et al., 2017). The authors report that the frequencies of "high-attainment alleles" increased in Europe since the time when Agamemnon led the Achaeans against Troy. For this kind of study, the main limitation is not variable genetic linkage between causal variants and associated SNPs, but the poor quality of the ancient genomes. In Woodley of Menie et al. (2017), only 99 ancient genomes were available and average coverage of the SNPs in the ancient genomes was only 30%. Correlations between currently available polygenic scores and the relevant outcomes are only about 0.2. Therefore in comparisons between groups, such as high school dropouts and university graduates, non-trivial sample sizes are required for statistically significant results. Despite the limitations of this inaugural study, whose results may or may not be confirmed by future work, the field of archaeogenetics is developing rapidly. If this first result holds up, it implies that Europe's progression from prehistory to classical antiquity, Dark Age, Renaissance and the scientific and industrial revolutions was driven in part by genetic selection. Unfortunately, archaeogenetic studies of this kind are difficult or impossible to perform in tropical climates, where fossil DNA is unlikely to survive for thousands of years.

Compared to archaeogenetics, studies that examine the direction and strength of genetic selection in living populations are low-hanging fruit. In the most detailed study of this kind so far, Kong et al. (2017) described that in Iceland, a polygenic score for educational attainment was associated with delayed reproduction and lower number of children, with the result that the average polygenic score declined by about 0.1 standard deviation during the last century. Interestingly, part of the antinatalist effect of this polygenic score was independent of educational attainment. This means, the cognitive abilities or personality traits that were affected by the causal genetic variants impaired reproduction even of those who did not stay in school. In modern societies, simply closing down all schools would not eliminate selection against higher cognitive ability or the personality traits needed for success in school. Findings like those of Kong et al. (2017) are not unique to Iceland. Similar results have been obtained in smaller studies in the United States (Beauchamp, 2016; Conley et al., 2016). Therefore the claim that the subfertility of highly educated and intelligent people in modern societies leads to "genetic deterioration" (Lynn, 2011) is becoming widely accepted within science, although scientists may prefer a more neutral term (perhaps "genetic realignment"?) to describe their results to the public. Researchers are now beginning to study the question the proper way: one gene at a time. But are the researchers' polygenic scores high enough to finish the job, and will non-scientists' polygenic scores be high enough to understand what the results imply?

References

Beauchamp, J.P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences USA* 113: 7774-7779.

Belsky, D.W., Moffitt, T.E., Corcoran, D.L., Domingue, B., Harrington, H., Hogan, S., … & Caspi, A. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science* 27: 957-972.

Conley, D., Laidley, T., Belsky, D.W., Fletcher, J.M., Boardman, J.D. & Domingue, B.W. (2016). Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proceedings of the National Academy of Sciences USA* 113: 6647-6652.

Domingue, B.W., Belsky, D.W., Conley, D., Harris, K.M. & Boardman, J.D. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA Open* 1(3): 1-13.

Gunnarsson, B., Jónsdóttir, G.A., Björnsdóttir, G., Konte, B., Sulem, P., Kristmundsdóttir, S., … & Stefánsson, K. (2016). A sequence variant associating with educational attainment also affects childhood cognition. *Scientific Reports* 6: 36189.

Kong, A., Frigge, M.L., Thorleifsson, G., Stefansson, H., Young, A.I., Zink, F., … & Stefansson, K. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences USA* 114: E727-E732.

Lynn, R. (2011). *Dysgenics: Genetic Deterioration in Modern Populations*. London: Ulster Institute.

Piffer, D. (2016). Evidence for recent polygenic selection on educational attainment inferred from GWAS hits. *Preprints* 2016110047 (doi: 10.20944/preprints201611.0047.v1).

Piffer, D. (23 May, 2017a). Evidence for recent polygenic selection on educational attainment and underlying cognitive abilities inferred from GWAS hits: A Monte Carlo simulation using random SNPs. *Preprints* 201701.0127.v3 (doi: 10.20944/preprints201701.0127.v3)

Piffer, D. (2 June, 2017b). New genes, same results: Group-level genotypic intelligence for 26 and 52 populations. https://topseudoscience.wordpress.com/2017/06/02/new-genes-same-results-group-level-genotypic-intelligence-for-26-and-52-populations/

Piffer, D. (8 June, 2017c). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using recent data. *Preprints* 2017060039 (doi: 10.20944/preprints201706.0039.v1)

Racimo, F., Berg, J.J. & Pickrell, J.K. (4 June, 2017). Detecting polygenic adaptation in admixture graphs. *bioRxiv* doi: http://dx.doi.org/10.1101/146043

Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., ... & Albrecht, E. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340: 1467-1471.

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R.I., Krapohl, E., … & Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, advance online publication.

Ware, E.B., Schmitz, L.L., Faul, J., Gard, A., Mitchell, C., Smith, J.A., … & Kardia, S.L.R. (2017). Heterogeneity in polygenic scores for common human traits. *bioRxiv* doi: http://dx.doi.org/10.1101/106062

Woodley of Menie, M.A., Younuskunja, S., Balan, B. & Piffer, D. (2017). Holocene selection for variants associated with cognitive ability: Comparing ancient and modern genomes. *bioRxiv* doi: http://dx.doi.org/10.1101/109678

Gerhard Meisenberg

MANKIND QUARTERLY 2017 57:4 487-518

Decapitation of Humans and Anthropomorphic Figurines in the Kura-Araks Culture from Armenia

Anahit Yu. Khudaverdyan*
National Academy of Science, Republic of Armenia

*Address for correspondence: Institute of Archaeology and Ethnography, National Academy of Science, Republic of Armenia, Yerevan, 0025, Charents st.15, Armenia, e-mail: ankhudaverdyan@gmail.com.

Anthropologists have always been interested in ritualistic behaviors and cultural practices of our ancestors, including the people of the Early Bronze Age from Armenia. In this article we present the analysis of human skeletons at Landjik site in northern Armenia during the period of the Early Bronze Age. During the 1996 excavation of the Landjik site on the Shirak Plateau, a burial site with the skeletal remains of 8 adults and 2 infants was found. In 1996, a study of the remains was initiated at the Institute of Archaeology and Ethnography, National Academy of Science, Republic of Armenia (Khudaverdyan, 1996). A more detailed study of the remains revealed the presence of cutmarks on bones of three of the adults. Distinct cut marks on the occipital bone suggest decapitation of the living individual. Also, headless figurines from the Early Bronze Age, found on the territory of the Armenian highlands, hint at this practice of head removal.

Key Words: Armenia, Kura-Araks culture, Skeletal analysis, Decapitation, Anthropomorphic figurines.

The Kura-Araks culture or Shengavitian still is one of the most enigmatic and poorly documented phases of cultural and social development in the prehistory of the Southern Caucasus. According to the few ^{14}C datings available so far, it originated around the middle of the 4th millennium BC. The economy was based on farming and livestock-raising, especially of cattle and sheep. The people grew grain and orchard crops. They raised cattle, sheep, goats, dogs, and in later phases, horses. Horse bones became widespread beginning about 3,300 BCE, with signs of domestication (Anthony, 2010). There is evidence of trade with Mesopotamia as well as Asia Minor (Avetisyan, 2015). The creators of the Kura-Araks culture are considered to be indigenous to the Armenian Highland and Transcaucasia (Khudaverdyan, 2011).

This paper gathers bioarchaeological information on decapitated individuals and anthropomorphic statues from the Armenian Highland and offers preliminary conclusions on the range of their possible functions and meanings. Almost all theoretical and research approaches to violence begin with the assumption that, at its core, violence represents the breakdown of meaning, the advent of the irrational, and the infliction of physical harm. Through such work, we may gain a better understanding of past human conceptual systems, especially as they relate to ritual practice, embodiment and symbolism.

Separating the Head from the Body as Cultural Practice

Bodily dismemberment – particularly human decapitation – is a powerful statement that surfaces in the iconography, mortuary practices, and political agendas of many cultures. This form of violence is evident in almost every time period and every region of the world (e.g. Andrushko & Torres, 2011; Brickley & Smith, 2006; Judd, 2004; Riches, 1986; Walker, 2001), and often those acts of violence are associated with issues of economics, territory, morality, or religion. Hoards of skulls, images of decapitated individuals, and depictions of disembodied heads occur along a broad spectrum of ancient contexts, ranging from Mayan skull platforms and head ornaments (Moser, 1973) to displays of bloodied heads in the Roman forum (Hinard, 1985; Lintott, 1968; Richlin, 1999). In each case, the headless bodies and bodiless heads can encapsulate specific, though not necessarily identical, meanings. Some limited examples appear in the artistic traditions and burial practices of Anatolia and Greece (Talalay, 2004). Very little research, however, has focused on the notion of decapitation throughout the prehistoric Caucasus.

From the literature (Hill, 2006) we can define the six functions of the human head or skull identified below: apotropaic, synecdoche, cathexis, mnemonic device, object of forgetting, and dedicatory offering. These functions are not mutually exclusive. In other words, human heads may be used in multiple ways by a single cultural group.

Apotropaic

The human head or skull has apotropaic properties: the ability to ward off evil or misfortune. The decapitated Chinese war-god Guandi appears in a dream to a monk, demanding his head. Upon being reminded by the monk that he decapitated many others, the warrior becomes a guardian and defender of the Buddhist faith (Duara, 1988). Other individuals also benefit from protection conferred by the head. The Vasyugan shaman enjoys the protection of the "spirit of the head" on ecstatic, trance-induced journeys (Eliade, 1964, p. 89). Allen (2002) recorded among Quechua speakers of Sonqo (Peru) that a skull placed in a niche of a storeroom was believed to protect the contents. A related belief is that the body of the deceased might harbor dangerous forces and, through appropriate treatment such as decapitation, the danger or threat may be averted (Todd, 1987). The deposition of human skulls has been interpreted as evidence for a headhunting cult, perhaps relating to enemy dead, whose deposition could provide symbolic protection (Wilson, 1981). The treatment of the head among the Jívaro of the lowlands of eastern Ecuador, documented in the 1950s, may be considered apotropaic; the heads of enemies were known to contain an avenging soul (muisak) that must be properly released in order to prevent death or mishap. The Celts sometimes took the heads of the slain and placed them on their houses so as to receive protection from their enemies' ghosts (Onians, 1973).The ancient Egyptians were also known to mutilate their enemies and to exhibit the separated body parts. Two rows of decapitated individuals, with their severed heads tucked between their legs, are depicted on the reverse side of the famous Palette of Narmer, which dates back to ca. 3,000 BC (Stokstad, 2002). Proper treatment involved decapitation of the enemy, curation, and rubbing charcoal on the head in order to blind the soul inside (Harner, 1962; Taylor, 1993).

Synecdoche

The idea of the trophy head is generally consistent with the concept of synecdoche, representing the slain enemy when his skull is displayed in order to demonstrate prowess in warfare. Examples include the Ohio Hopewell (50 BC–AD 350) of eastern North America (Seeman, 1988) and the Iron Age Celts of Europe (450–100 BC). Petres (1972) has interpreted deposits of the heads of warrior-age males from the Iron Age in Hungary as synecdochical; that is, the male head represents the entire warrior. Many of the Nasca heads from ancient Peru were modified by enlarging the foramen magnum, perforating the frontal bone, inserting a wooden toggle and attaching a rope for suspension (Proulx, 2001; Williams et al., 2001), presumably for display. A synecdochical function is also consistent with the use of the skull or head of the deceased as an object of veneration (Urcid, 1994). In situations in which the preservation or maintenance of the deceased's body in toto is impossible or inconvenient, the skull, one of the most identifiable skeletal elements in the human body (Becker, 1988), may be retrieved after death and venerated in place of the entire person. Modified human skulls and mandibles recovered on Kodiak Island and in southeast Alaska have been interpreted as trophy heads (de Laguna, 1933).

During the fifth century BC, the Greek historian Herodotus graphically detailed how Central Asian Scythian warriors sought and preserved human trophies. “The Scythian warrior must bring the king the heads of all those he had killed in battle. A cut is made in the head near the ears, and then the head is taken by the hair and shaken out of the skin” (Herodotus, cited in Borodovsky & Tabarev, 2005, p. 87). They then proceeded to remove the skin from these heads, thus creating fleshy “handkerchiefs.” These handkerchiefs were then hung on the bridles of horses belonging to warriors as symbols of their victory in war. Herodotus goes on to report that drinking cups were fashioned from the remaining skulls (Burton, 1864).

In the case of the ritual sacrifices that took place in Armenia during the Late Bronze Age and Early Iron Age it has been generally assumed that appeasing the gods was the main purpose of such bloody rituals (Khudaverdyan, 2014b). Two females from the Sevan region (Noraduz, Karmir) show evidence of decapitation. The evidence for the removal of heads and scalping in Armenia comes from the Shirakavan site dated to the Late Iron Age. One of the most interesting and unique discoveries made at Shirakavan are the detached human skulls (Khudaverdyan, 2014a). The skulls of the individuals, both male and female and ranging in age from 16 to 55 years, had been detached from their bodies and placed into burials in the Shirakavan site. Of the detached human skulls, two are females and one is male. The skulls are not complete because their mandibles are missing. Teeth were extracted from the skulls and buried separately. Skulls were buried with other human remains (Figure1). Acephalous skeletons were not discovered in the cemetery. We can say without a doubt that the detached human skull is not reserved for individuals of a specific gender or age. All the skulls were found together with various objects that can be interpreted as ritual offerings. The skulls were not thrown carelessly into burials but carefully placed into them. Human heads were clearly part of an elaborate ritual complex within Shirakavan society. Since the detached skulls belong to young females and one adult male, one can interpret these as human sacrifices rather than venerated ancestors: sacrifice as cosmological transition, ritual being associated with social differentiation (Khudaverdyan, 2014a). Further references to trophy taking can be found in Late Antiquityat Beniamin site (Khudaverdyan, 2015a).

Figure 1. *The arrow indicates the skull-trophy (Shirakavan, burial 1).*

Smith (1997), for example, has suggested that human trophy taking as a prestige-motivated activity may prove to be an important variable in documenting incipient social complexity because it may be linked to differential mortuary treatment. At the American Northwest Coast, for example, burial practices changed significantly from midden burials in the period AD 500–1000, which led Ames and Maschner (1999) to speculate that the coast-wide change in burial practice may have been associated with other changes in social complexity, which in turn coincided with the escalation in warfare.

The most infamous Biblical example of the display of a severed body part is found in the New Testament: After being delighted by a dance performed by Herodia’s daughter, Herod said to the girl, “Ask of me whatever you wish and I will grant it to you” (Mark 6:22). After consulting with her mother, the girl replied, "I want you to give me at once on a platter the head of John the Baptist" (Mark 6:24). Herod then, "promptly dispatched an executioner with orders to bring back his head. He went off and beheaded him in the prison. He brought in the head on a platter and gave it to the girl. The girl in turn gave it to her mother" (Mark 6:27-28).

Cathexis

A further function of the head is as a cathected object, such as a charm or talisman. Cathected objects are those believed to be imbued with supernatural power or energy. The relics of saints are examples of cathected objects. A bone fragment of a saint could cure illness (Schmandt-Besserat et al., 2010) or confer good fortune. Once identified as relics, the heads of saints were decorated, enshrined and visited by the faithful seeking cures or good fortune because the relics were believed to be efficacious (Bynum & Gerson, 1997). In some South American cultures, the head was believed to contain or embody special powers of benefit to the one who took the head or possessed it. During the nineteenth century, the Munduruku of the upper Tapajós River of eastern Brazil reportedly curated the heads of their enemies and used them as talismans to increase the availability of game animals (Murphy, 1959). Benoît (1975, p. 258) writes, "The severed human head is the most ancient talisman of mankind. Buried beneath the door-step of a house, enclosed in the rampart, or carved on the rampart of the fortress (as at les Baux, Tarragone, Albarez and Lugo in Celtiberia), it was intended to keep malevolent powers at a distance and terrify the enemy, like the Gorgon head on the shield of Athena."

Mnemonic device

The mnemonic roles of monuments, architecture and specific artifact classes are becoming increasingly familiar subjects of study in archaeology (e.g., various chapters in Van Dyke & Alcock, 2003; Williams, 2003a). The body as a site of memory is already well-established in European ethnography (e.g., Verdery, 1999) and medieval archaeology (e.g., Williams, 2001, 2003b). The classic example of the head as a memento of an ancestor are the plastered Neolithic skulls of the Near East (7,200–6,000 BC for the Levant). Kenyon (1957) was the first to propose that these skulls served as "mementos," although recent research indicates that skulls fulfilled multiple functions in ancient Near Eastern mortuary ritual, including their possible use as fertility or apotropaic devices (Bonogofsky, 2001, 2003, 2004; Pearson, 1999). Decapitated Nasca heads are another example of the possible use of heads as mnemonic devices to facilitate ancestor worship. Pointing to the well-documented practice of preserving the bodies of dead ancestors among the Inca and care-taking of the dead by modern Quechua speakers in the Andes (Allen, 2002), researchers such as Carmichael have postulated that the Nasca employed trophy heads as "generic reference[s] to the collective ancestors and their life-giving powers" (Carmichael, 1994, p. 84). The curation of the head, then, reminded the keeper of the central role of the ancestors in maintaining the community.

Object of forgetting

A penultimate function is the inverse of a mnemonic device; that is, the head is manipulated so as to create an object with which to forget or erase the social identity or memory of the deceased. Among many twentieth-century native Amazonians, the physical modification of the corpse is a prerequisite to forgetting (Oakdale, 2001; Taylor, 1993). The Wari' of western Brazil, for example, explain their dismemberment and cannibalism of the deceased as a service to the family, as a way to remove the "focus of remembering" (Conklin, 2001) and obliterate the body as a tangible link between the deceased and surviving kin. Medieval citizens feared individuals with leprosy out of uncertainty, misinformation, self-preservation, and ignorance. By virtue of being considered evil outcasts, people with the disease were symbolic representations of evil. Disabled individuals were treated as second class human beings who were unfinished and deformed due to a sin or cheat of nature. In Medieval Armenia, there is clear archaeological evidence for the decapitation of three individuals from Dvin (Khudaverdyan, 2015b). The victims were sometimes savagely beaten and tortured before their heads were cut off. Cranial sharp-force weapon injuries were observed, which varied from small isolated punctures to multiple linear injuries. These unfortunate individuals were expelled from society and lived in marginal conditions.

Duncan and Schwarz (2015) explore spatial distribution of skeletal remains in a postclassic (AD 950–1524) Maya mass grave using Ripley's function to show that bodies were intentionally fragmented and manipulated, on the basis of side and element in a commingled secondary context. This cursory overview of the myriad forms of displaying severed human skulls is not meant to be exhaustive as there are countless other examples and numerous other reasons for engaging in these behaviors.

Symbolic Decapitation

Analyzed samples of human (anthropomorphic) figurines from the Kura-Araks culture on the territory of the Armenian highlands show a greater number of fragmented figurines than whole figurines (Figure 2). All of the studied figurines were found in different parts of the settlements, never in graves. The figurines appear to have been mutilated and damaged, the most common feature being removal of the head. The symbolic intent of these headless anthropomorphic figurines certainly had social ramifications. The fact that headless anthropomorphic figurines appeared in settlements suggests the symbolic and social significance of heads during the Early Bronze Age, particularly as part of a shared system of ritual practices. We can propose that, at some level, the objectified images of headless bodies and decapitated remains served to mediate the complex relationships between the living and the dead (Last, 1998).

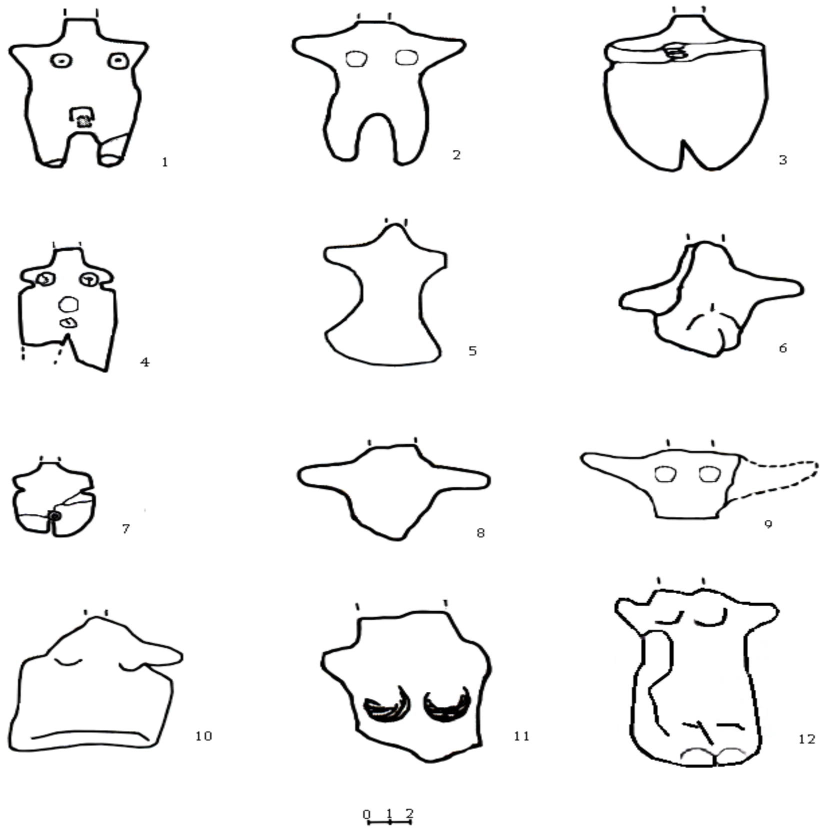

Figure 2. *Figurines 4th millennium BC on the territory of the Armenian highland: 1, 3, 4, 6, 7 – Agarak in the southern foothills of Mt. Aragats on the western bank of the Amberd; 2 – Shengavit (Yerevan, located on a hill south-east of Lake Yerevan); 5 – Harich (Shirak Province); 8, 10 – Mokhrablur (Echmiadzin); 9 – Jrahovit (Ararat region); 11, 12 – Norsuntepe (modern eastern Turkey).*

Anthropomorphic figurines are recovered in a variety of archaeological contexts, many of which can be clearly defined as domestic, burial, and ritual. Throughout the figurine corpus in the Armenian highlands, and throughout the Chalcolithic and Early Bronze Age, just over half of the figurines appeared to be female or to have exclusively female characteristics sufficient to lead us to conclude that they are probably female. A very small percentage were obviously or probably male (Figure 2: Agarak 7). The mother goddess explanation is one of many possible explanations for figurines in the Neolithic (Malone, 1998; Mellaart, 1967). Ucko (1962, 1968) looked for ethnographic analogs and countered the mother goddess hypothesis with suggestions that Neolithic figurines filled a variety of roles which included rituals for curing, protection, initiation and marriage, as well as to support oral narratives. Biehl (1996) concluded that the figurines of the Gradesniča-Krivodol culture complex (European Neolithic) represented individuals within the society rather than a given deity. These individuals, he suggested, were being represented through figurines as they participated in “life's most pivotal moments: birth, the need for food, and death” (Biehl, 1996, pp. 170-171). The last stage of a figurine's creation, Biehl notes, was often its destruction.

The Landjik site (Armenia, Shirak Plain, Figure 3) was characterized by unusual findings with regard to the human skeletons buried there; these were probably victims of ritual sacrifice. The goal of this study is to determine, as far as possible, the relationship between headless anthropomorphic figurines and human decapitation among the tribes of the 4th millennium BC Kura-Araks culture.

Materials and Methods

The Landjik archaeological site from the Early Bronze Age, dated by means of grave pottery, was discovered during field work in the Shirak Plain (Armenia) in 1996. The human skulls analyzed for this article were excavated by archaeologist Levon Petrosyan. The Landjik site, located 1800 m above sea level in the Lesser Caucasus, includes one burial mound (Figure 4). It is a slight elevation located in the eastern part of the village of Landjik, containing isolated human and animal bones, 10 ceramic vessels, ceramic fragments, and stones. Like at other Early Bronze sites (Kaps, Keti), people of Landjik buried their deceased in mass burials, in which skeletal remains were rarely found in their correct anatomical position. Archaeologists believe that these remains were grouped according to special ties, such as kinship (Petrosyan, 1996). Only one skeleton was preserved in anatomical order. Postcranial bones of individuals (after skeletonization) were collected and stacked below the skulls. These characteristics result from postmortem manipulation of the remains. Ceramic vessels, located on the floor inside the burial chamber, would typically contain food or drink.

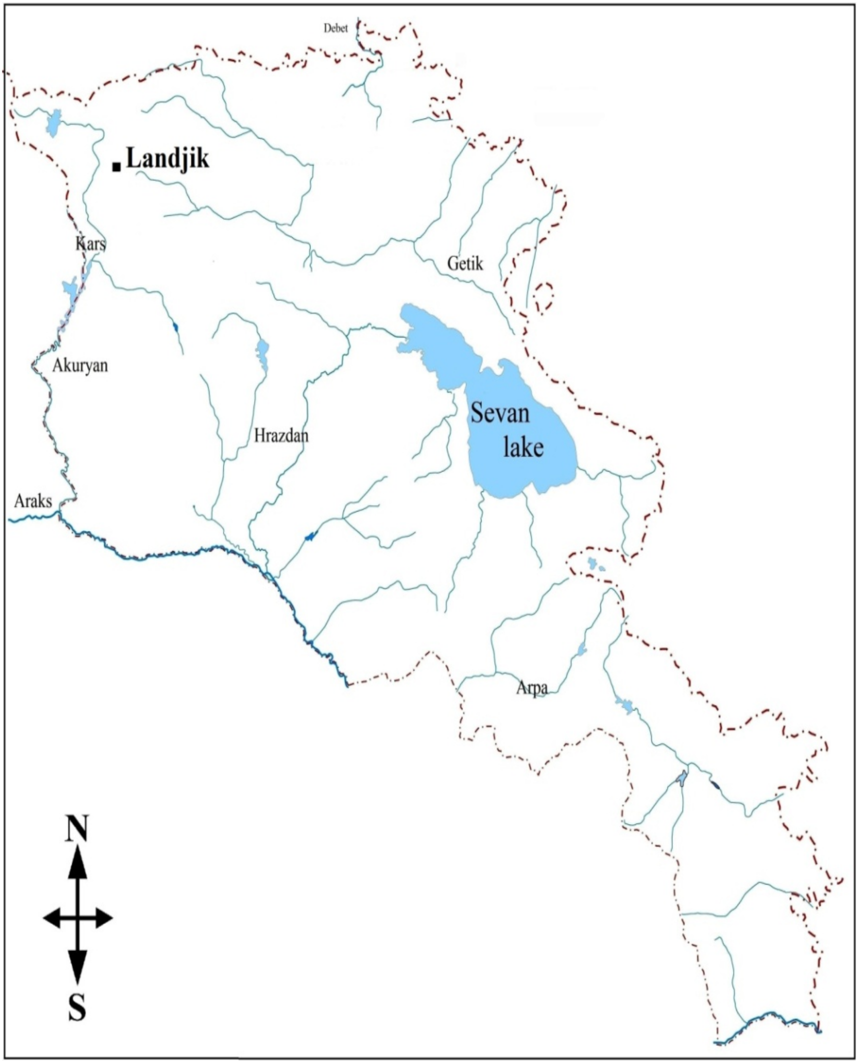

Figure 3. *Map of Armenia, showing the location of Landjik.*

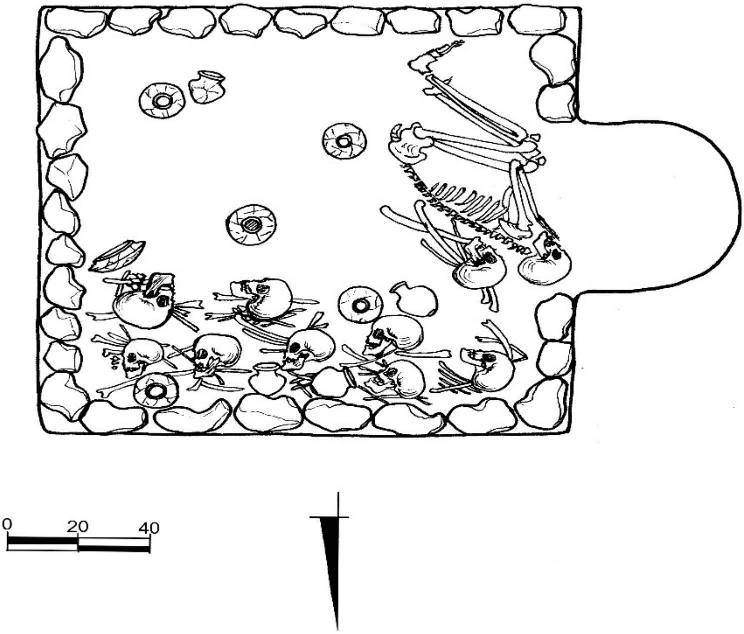

Figure 4. *Plan of the burial (artist Ani Sahakyan).*

Most of the skulls were in a satisfactory state of preservation, suitable for estimation of sex, age, and the pathological observations that are the main subject of this work. Unfortunately, no post-cranial bones have been collected. Sex and age at death of the individual were determined according to standard osteological methods. Morphological features of the crania were used for the identification of sex (Brothwell, 1981). The degree of cranial suture closure (Meindl et al., 1985) was used for adult age estimation. Measurements were taken as outlined in Alekseev & Debets (1964). Non-metric traits have been recorded for these skulls (Movsesyan et al., 1975) and dentition (Zubov, 1968) in order to allow future comparisons with findings from other sites of Armenia.

Gross observations of abnormal changes appearing in ancient skeletons provide the basic information for paleopathological diagnosis and interpretations of perimortem trauma and manipulation. In the present study, pathological changes are completely described and given tentative diagnoses.

By means of an osteological analysis, it was determined that a total of 10 were present in this tomb context: 2 children, 6 women and 2 men. Preliminary results for Landjik have been published previously (Khudaverdyan, 1996;

Khudaverdyan et al., 2015). Eight out of 10 individuals from Landjik showed evidence of cribra orbitalia. The mild forms of this stress indicator were observed in all age and sex cohorts. Dental caries was recorded on the teeth of three adult individuals (3/8). Enamel hypoplasias affected 5 individuals (N=10). Three adult individuals met the criteria for antemortem tooth loss. Maxillary teeth were more often affected than the mandibular dentition. One out of 8 observable dentitions exhibited some form of alveolar bone loss (№ 1: M2). Calculus was recorded on the teeth of 5 individuals. Button osteomas were found on the frontal and parietal bones of four skulls. This study follows up on the previously obtained results from the Landjik site and presents new results derived from a detailed analysis of the finds, in particular those suggesting decapitation.

Traumatic injuries to the skeleton can occur in many distinct patterns depending on the force of impact, its direction, and the structural integrity of the affected bone (Galloway, 1999; Lovell, 1997). The bodies of multiple individuals showed evidence of possible perimortem trauma to the head. Perimortem injuries refer to any injuries sustained while the bone was still fresh and malleable. There are certain patterns that can be used to distinguish perimortem from postmortem. Due to the plastic nature of bone, perimortem fractures will often exhibit beveling of the surface outward in the direction of the force applied, especially on the laminar bones of the cranial vault, which will exhibit linear, comminuted, or puncture fractures with similar beveling (Lovell, 1997). The edges of a perimortem fracture will also appear irregular since fresh bone will flex and pieces will remain attached (Sauer, 1998). Fractures of the cranial vault, depending on the force of impact, will include radiating and possibly concentric fractures that will stop upon reaching an open suture or fracture line (Berryman & Haun, 1996).

Injuries of the skull base can be divided into bursting fractures and bending fractures. Bending fractures are caused by direct, focused trauma to the skull. Depression of the bone at the site of impact results, which typically produces comminuted or perforated fractures. Bursting fractures are caused by objects with a wide surface area and indirect trauma to the cranial bones. The resulting force is transmitted and, in areas where the bone is thin, minimal elasticity results in breakage, i.e., bursting fractures.

When decapitation was encountered, a full description of the decapitation was completed. This included which bones were affected, the direction of the blow, any indication of what weapon may have been employed, and a general description of the individual cut marks associated with the decapitation. This information was expanded upon with a full schematic drawing of the location of the decapitation related wounds and a photographic record to fully record the trauma. Damage to the upper cervical vertebrae (and also C7 or T1 in some instances), mastoid processes, occipital regions, the posterior parts of mandibles and first ribs have been considered as good markers of decapitation (Anderson, 2001; Ardagna et al., 2005; Aufderheide & Rodríguez-Martín, 1998; Buckberry & Hadley, 2007; Khudaverdyan, 2015b). Beheading-related traumas were also observed on the odontoid peg (McKinley, 1993) and transverse processes of vertebrae when an axe rather than a sword has been used (Waldron, 1996). Even if no evidence has been left on the bones, some aspects of the burial context can be indicative of decapitation such as the absence of a head (although bones can eventually be destroyed or lost through post-depositional processes such as intrusive burials, animal activities, and environmental conditions) (Okumura & Eggers, 2008), the presence of a head without other postcranial elements (Nagaoka & Abe, 2007), or the placement of a head in a nonanatomical position (Boylston et al., 2000).

Results

A wealth of information about the life of an individual can be gleaned from examination of skeletal remains. “Death is the end result of an accumulated set of biological, behavioral, and cultural responses to challenges in the social and physical environment” (Martin & Akins, 2001, p. 223). The physical condition of a person’s body — indications of overall health, nutrition, disease processes, injury — can reveal information about that individual’s role in life. Furthermore, analysis of the manner in which an individual’s body was treated at death can help to enrich the understanding of mortuary customs and their cultural significance, contributing to a more nuanced understanding of mortuary ritual.

There are two females from Landjik with evidence of possible decapitation: skeleton 3 aged 45 to 49 years, and skeleton 4 aged 50 to 59 years. There is also a male, skeleton 1 aged 60 to 65 years, who also displays possible evidence of decapitation. Thus all three were between 45 and 65 years old at the time of death.

Individual 1

The skull of a senile male, 60-65 years, was designated as Individual 1. Craniometric data are provided in Table 1. The skull has very large longitudinal and very small transverse diameter, dolichocranic, with a high vault. The forehead is medium wide, face is of small height. Nasal bones are small in height and medium in width. The following traits were present: *os wormii suturae squamosum, os wormii suturae lambdoidea, foramina mastoidea, canalis craniopharyngeus, canalis condyloideus, hypocone (UM2).*

Table 1. *Some anthropometric measurements in mm.*

Martin no.	Characters and indices	Individual 1 ♂ 60+ yrs.	Individual 2 ♀ 40+ yrs.	Individual 3 ♀ 50+ yrs.
1	Maximum cranial length (g-op)	208	183	180
8	Maximum cranial length (eu-eu)	133	134	122
17	Cranial length (ba-b)	139	131.5	125.5
9	Minimum frontal breadth (ft-ft)	97.5	90.5	93
12	Occipital breadth	111	114.5	96
48	Upper facial height	68	-	-
43	Upper facial breadth (fmt-fmt)	86	-	-
60	Maxillary alveolar length (incision-alv)	59.8	-	-
61	Maxillo-alveolar breadth (ecm-ecm)	61.5	-	-
62	Palatal length (st-o)	50	-	-
63	Palatal breadth between the second molars (enm-enm)	31	-	-
55	Nasal height (n-ns)	49.9	-	-
54	Nasal breadth (al-al)	25	-	-
51	Orbital breadth (d-ec)	39	-	-
51a	Orbital breadth (ect-d)	36	-	-
52	Orbital height Bicondylar width	30.5	-	-
8:1	Cranial index	63.95	73.23	67.78
17:1	Height-length index	66.83	71.86	69.73
17:8	Height-breadth index	104.52	98.14	102.87
9:8	Fronto-transverse index	73.31	67.54	76.23
54:55	Nasal index	50.11	-	-
52:51	Orbital index (mf)	78.21	-	-
52:51a	Orbital index (d)	84.73	-	-
63:62	Palatal index	62	-	-

This cranium exhibits antemortem tooth loss of the maxillary right M2 and M1. Determining the etiology of antemortem tooth loss is difficult as evidence may have been lost, especially in instances of carious teeth (Hartnady & Rose, 1991). However, the close association between periodontal disease, dental caries, and antemortem tooth loss is well established, especially in archaeological populations (Larsen, 1997). Resorption of alveolar hard tissues was moderate in the maxilla (M2).

Excavation and analysis of the human remains revealed some evidence of dismemberment associated with this individual. On the inferior aspect of the occipital bone, there is a probable peri-mortem sharp-force weapon injury. Widespread damage was identified on the structures of the base of the skull, including the edge of the foramen magnum, the mastoid process, and the inferior surface of the occipital bone. In the base of this individual cranium mechanical breaks of occipital condyles and damage of the mastoidal were noted (Figure 5). This bone damage was inferred to have occurred around the moment of death. These injuries are extremely similar to other published cases: beheading of the person who is in a vertical position (Khudaverdyan, 2015b; Manchester, 1983). Specific destructions of occipital condyles, say a linear break of a mastoidal, indicate that a strong enough blow at the skull was struck from behind. Holding the victim by his hair, the head was cut off with a sword (Figure 6). This could have been done as a means of murder or execution; it could have been accomplished, for example, with an axe, sword, or by other means.

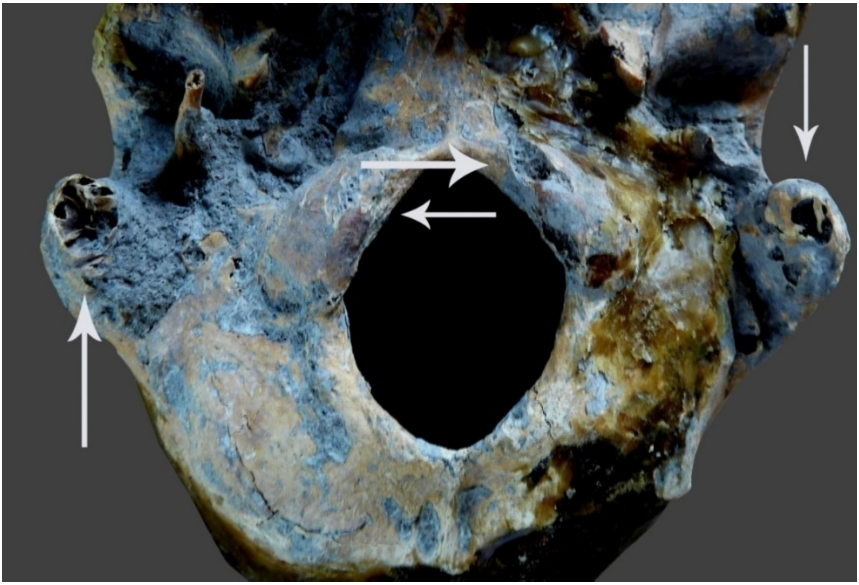

Figure 5*. Possible perimortem breaks of occipital condyles and damage of the left mastoidal (individual 1).*

As we know, both humans and animals may be victims of sacrificial bloodletting and appear to have been participants in these rites (Green, 1998). Such rites normally have targeted the throat and carotid arteries and jugular veins. These structures can be severed just below and behind the ear (e.g. at the pressure point), but would not require removing the head or any contact with the underlying vertebrae. Given the location of the pressure point, one might expect cutmark evidence in the region of the mastoid process (Hamilton, 2005), if at all. Importantly, near-contemporary animal sacrifices appear to have been performed using a blow to the head. Perhaps a bloodletting ritual was responsible for these decapitations.

Figure 6. *Reconstruction of the human decapitation (artist Ani Sahakyan.)*

Knight (1991, p. 214) records the tell-tale signs of the cut throat: "The classic description of the cut throat is of the incisions starting high on the left side of the neck below the angle of the jaw, which pass obliquely across the front of the neck to end at a lower level on the right." This is similar to the pattern noted at Kempston (Boylston et al., 2000). There are other indications which may indicate beheading. Strong enough injury of the skull base caused a fracture in the bones at the base of the skull. A fissure was seen extending from the side of the foramen magnum up linea nuchae inferior (left). A transverse fracture is located on the right side in the area of the linea nuchae inferior. This appears to be pathological rather than taphonomic. The injury appears to have resulted from a single blow to the back of the neck.

Individual 2

The second possible victim is female, 40+ years. The skeletal remains of this individual were poorly preserved, almost all facial bones were absent or fragmented. The skull is gracile in dimensions, the cortices are thin. Macrorelief on the skull is weak. The skull is dolichocranic, with a high vault and with very large longitudinal and middle transverse diameters. The forehead is narrow in the front. The following traits were present: *foramina supraorbitalia, os wormii suturae squamosum, os wormii suturae lambdoidea, foramina mastoidea, canalis condyloideus, hypocone (UM2)*.

Asymmetrical articular facets are present on the inferior aspects of the occipital condyles of this skull. There were differences between the results on the right and left sides. The foramen magnum was considerably narrowed and asymmetrical. Exostosis and porosity in the ear canal are present in this female. Exostosis formations in the ear canal are regarded as an adverse effect of diving in cold water (Kennedy, 1986; Manzi et al., 1991). Exposure to wind or cold water causes the bone surrounding the ear canal to thicken and constrict the ear canal, sometimes to the point of complete blockage (known as occlusion). Porosity in the auditory or cribra, like lesions in the external auditory meati, are a side effect of exposure to a windy climate (Khudaverdyan, 2016).

A small circular button (or "ivory") osteoma on the outer table, presenting as a smooth bump not bigger than 4.2 mm in diameter, is present on the frontal bone (Figure 7). A button osteoma is a benign, slow-growing tumor which occurs almost exclusively in the skull (Aufderheide & Rodriguez-Martin, 1998; Ortner, 2003).

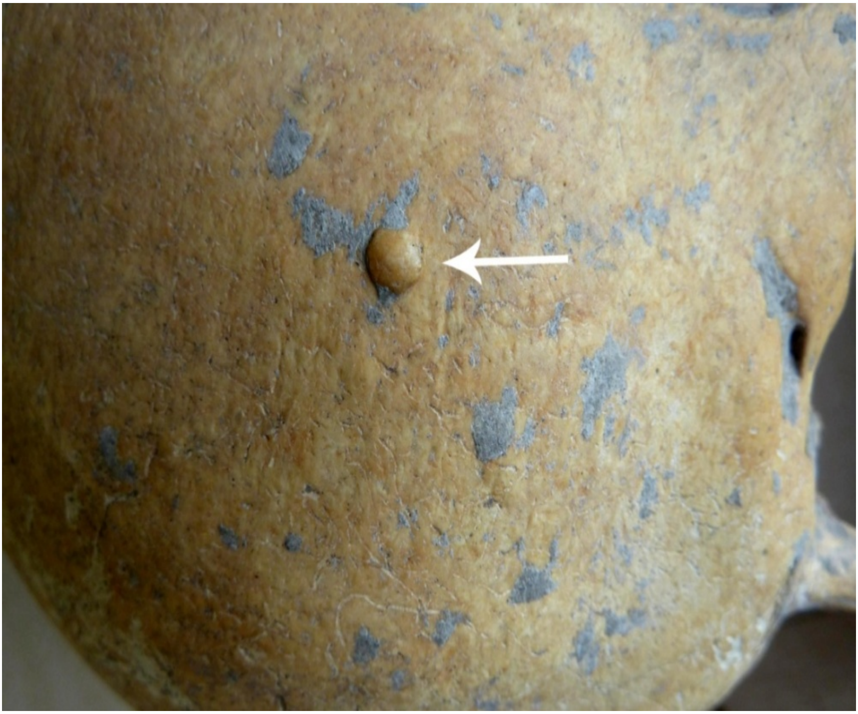

Figure 7. *"Button" osteoma on the frontal bone of individual 2.*

In the base of the individual's skull is mechanical damage of occipital condyles (Figure 8). Mechanical damage occurs when cartilage production, and thus material strength, is reduced, leading to cartilage fragmentation and fibrillation (Jasin, 2005). No cutmarks are present, but the condyles were severely eroded. However, in the absence of cut marks in the cranium and associated cervical vertebrae, it is not possible to determine if this is indeed a case of decapitation or the result of natural taphonomic processes. The microarchitecture of bone tissue clearly influences its mechanical properties, porosity and, ultimately, its resistance to post-mortem degradation. There is a healing porous lesion on the occipital bone that is probably a healing infection. The infectious process seems to be reflected on the occipital bone.

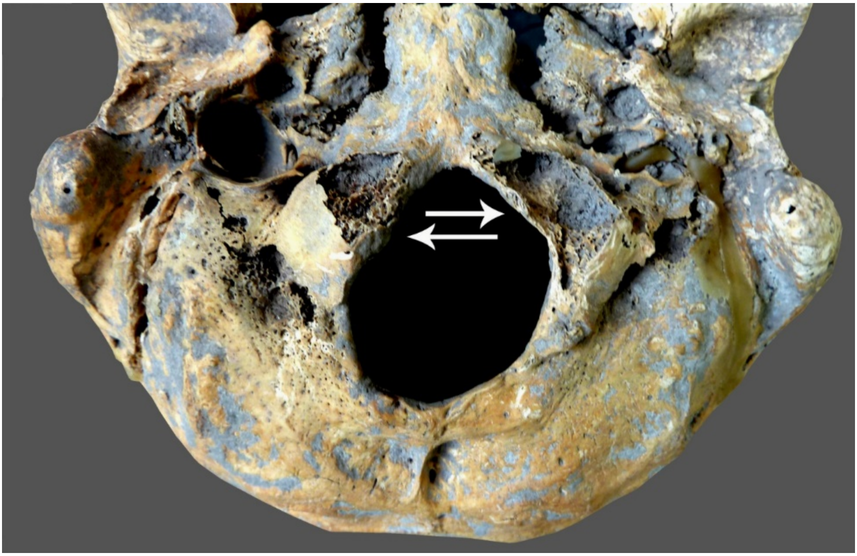

Figure 8. *Mechanical damage of occipital condyles, individual 2.*

Individual 3

The third individual is an adult female approximately 50-59 years. Almost all facial bones and the postcranial skeleton were absent or fragmented. The skeleton is gracile in dimensions, the cortices are thin. Macrorelief on the skull is weak. The skull has a large longitudinal and very small transverse diameter, dolichocranic with a small vault, the forehead medium wide. The following traits were present: *foramina supraorbitalia, os wormii suturae squamosum, os wormii suturae lambdoidea, foramina mastoidea, canalis condyloideus*.

A pathological analysis of this individual identified healed porotic hyperostosis of the cranial vault (temporal bone and pterional).This skeletal lesion has long been considered indicative of an anemia (Aufderheide & Rodriguez-Martin, 1998), osteomyelitis, inflammatory processes of the scalp (Larsen, 1997; Ortner, 2003), or nutritional deficiency, but current interpretations recognize that there may be many different causes of this condition.

This individual exhibits an antemortem well healed circular fracture of the right parietal bone (40×41.5mm), the inner table is intact (Figure 9). The head is an excellent location of the body for observations of violent interpersonal trauma. The majority of fractures occurring on the cranium are a result of direct trauma (e.g., a blow to the head). The cranium is the most likely area to be attacked using blunt weapons during individual, face to face conflict, making it the preferred skeletal element for studies relating to interpersonal violence (Galloway, 1999; Judd, 2004, 2006; Lovell, 1997).

Analysis of the cranium revealed some possible evidence of dismemberment. Mechanical breaks of right occipital condyle and oblique break of the right mastoidal were noted in the base of the individual's cranium (Figure 10). These fractures appear to be perimortem. It is likely that it was cut in the direction from right to left, angled. Perhaps, a bloodletting ritual was responsible for this apparent decapitation.

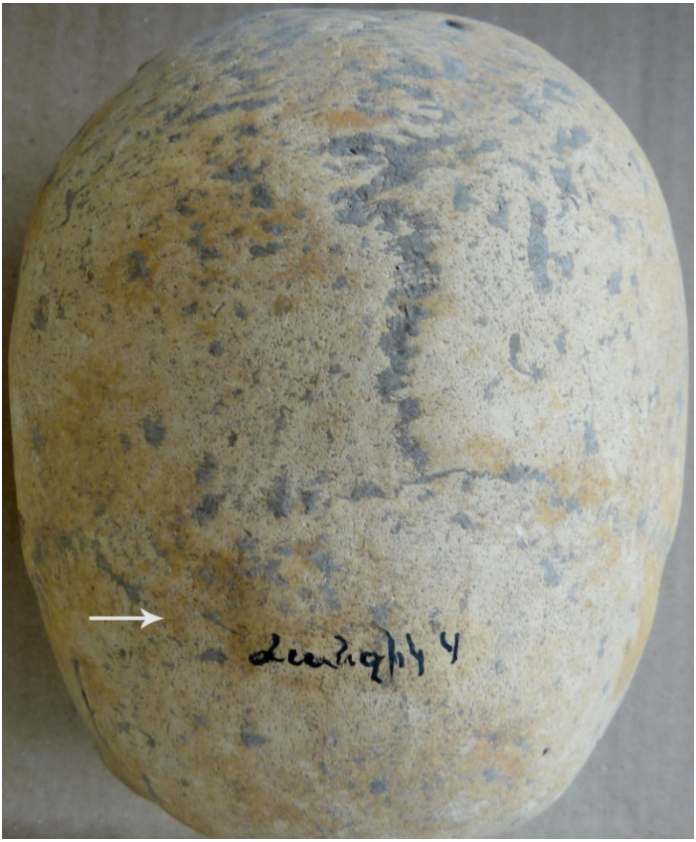

Figure 9. *Antemortem fracture of the right side of the frontal bone (individual 3).*

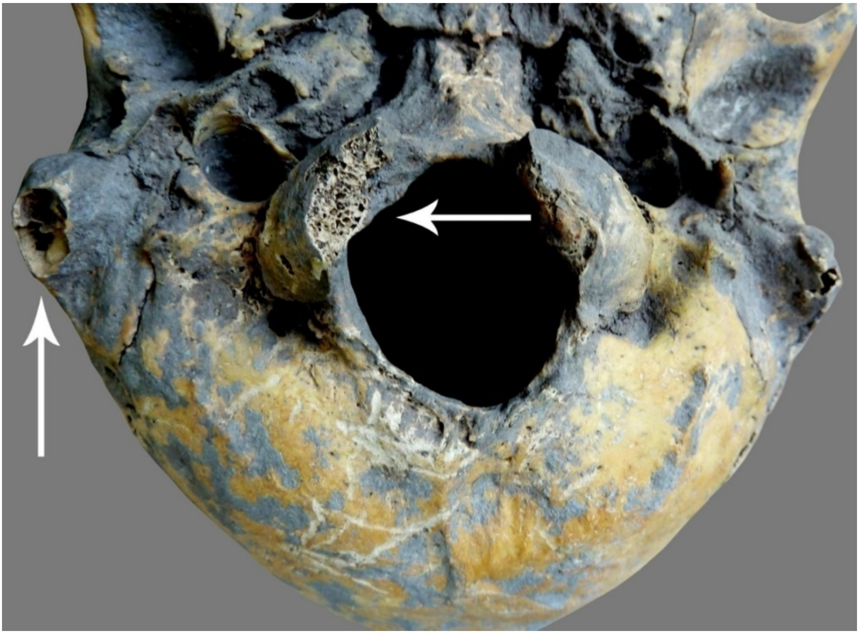

Figure 10. *Possible perimortem breaks of occipital condyles and damage of the right mastoidal (individual 3).*

Although counts of cut marks can be subjective, it is probable that a knife is the implement responsible for the perimortem trauma observed in this skull. No knives or cutting implements were found at the site, but an analogy can be made with Hamilton's (2005) data, suggesting these cuts were likely made by a long, beveled-edge, metal instrument. We know that crescent-shaped metal blade knives were associated with sacrificial activities and severed heads (Khudaverdyan et al., 2016, in press).

Discussion

The main objective of anthropology is furthering our understanding of human behavior across cultures in the past. Ritual is an active performance and a process of several stages which is both expressive and symbolic (Bell, 1992; Bloch, 1992). Ritual regulates and sustains boundaries, it maintains social structure and social order through symbolic representation, and it creates a feeling of solidarity among the participants. Ritual violence is challenging for several reasons: 1) the presence of ritual violence does not indicate a sacrifice took place; 2) what is considered violence and what is considered veneration differs depending on the context and culture (Duncan, 2005; Millaire, 2015); and 3) the presence of violent trauma does not necessarily indicate that a ritual took place (Eeckhout & Owens, 2008). For the identification of ritual violence (including sacrifice) in the bioarchaeological record (such as decapitation, throat-slitting, strangulation, bludgeoning, defleshing, dismemberment and trophy-taking), researchers have looked to a number of cues. The presence of fractures on skulls from Landjik has been suggested as representing intentional ritual violence.

Talalay (2004) and Bienert (1991) note that in the ethnographic record, skull trophies have been obtained after battle to underscore the importance of valor, bravery, pride and manhood, providing societies lacking hereditary rank with a system of social prestige. Thorpe (2003) explains the three main competing theories for warfare in archaeological contexts: territorial, reproductive, and status competition. In Landjik, neither drill holes nor enlargement of the foramen magnum were observed in the skulls (Boylston et al., 2000), nor evidence for any form of attachment such as the perforations made by nails in the crania from Entremont (Benoît, 1975). This makes it unlikely that these were trophy heads. Determining the identity of the decapitated individual can also contribute to understanding the broader cultural context in which decapitation practices are embedded. The status of the buried individuals as local group members or outsiders is another relevant point. If outsiders, they might in fact represent an enemy. The results of the non-metric morphological characters of teeth and cranial analysis show clear affinities between the individuals from Landjik and the series Kura-Araks from Armenia (Khudaverdyan, 2012, 2013).

As Hughes (1991) discusses in his book on human sacrifice in ancient Greece, ritual is usually formed out of a necessary basic function. For instance, the ritualized sobbing and dramatic displays of grief exhibited by some cultures during a funeral is a ritual that arose out of the basic human emotional need to vent their sorrow, go through the process of grieving, and move on with their lives. Similarly, rituals and ceremonies marking the beginning of adulthood arose out of a social need to have a definitive point distinguishing childhood from adulthood for legal reasons and other aspects of community life. This can probably be observed also in the ritual violence of the individuals from Landjik. Within the culture of the time appeasing the gods was most likely the main stated reason for conducting such "sacrifices," however there were probably many other more real and pragmatic reasons for inventing such rituals. Maybe they were prisoners of war or convicted criminals, witches, or victims of human sacrifice. However, it is important to note that the indication of a ritualized execution does not necessarily mean that a human sacrifice had taken place. A more detailed evaluation of this matter will depend on further work in the region.

On the Armenian Plateau, archaeologists have found numerous "decapitated" anthropomorphic clay figurines dating from the Early Bronze Age (or Kura-Araks culture) (Figure 2). All these figurines were found within settlements in different contexts (houses, pits, outside houses). It is very important to stress that figurines of the Kura-Araks culture are never found in burials. The head (and sometimes the neck) is missing, having been cut off probably with obsidian and other stone tools, and perhaps even polished after removal. The ancient art of Peru (including pottery and textiles) also frequently portrays scenes of decapitation or severed heads (Paul, 2001; Peters, 1991). In her review of iconographic evidence of ritual violence in the Andean region, Hamilton (2005), describes evidence of decapitation from the Initial Period (ca. 1500 BC) in the Casma Valley, Callejón de Huaylas and in the eastern highlands of Peru. The earliest evidence of ritual violence in the Andes comes from art from the Cupisnique culture (1500–1 BC) on the north coast of Peru. This art depicts a "decapitator" figure performing decapitation rituals.

In the opinion of Matić (2009), the practice of intentional fragmentation is reminding us that there is a connection between figurines and human bodies. The fact that human decapitation and decapitation of anthropomorphic figurines appeared in both burials and domestic settings illustrates the symbolic and social significance of heads during the Early Bronze Age, particularly as part of a shared system of ritual practices. Creating a figurine of a person and destroying it may have been done with the same purpose, to destroy the person represented by the figurine. We do not know, for example, whether the two phenomena were set in opposition to one another or whether they were thought of as interchangeable. The data are, unfortunately, too meager to let us parse these phenomena. Regardless of these uncertainties, it is apparent that decapitation of humans and anthropomorphic figurines in definite and specific venues or in association with other objects was not random. People make figurines and they are also created by them in the process of making, using, and breaking them. This then means that there is no clear difference between figurines as body re-presentations and the human bodies themselves, between original and alternative, between original and a "copy." Figurine as a body is actually a re-presentation of the idea of the body, or more precisely of what the body should be so that it can be fragmented. Human decapitation and decapitation of anthropomorphic figurines represent integrated acts of social memory: the reiteration of the naturalized order assists in both remembering and forgetting the dead. As physical actions, ritual and commemoration are linked to the production of shared memories and experiences in communities at different scales (see Connerton, 1989). In many cultural contexts commemoration and memory, especially when associated with death, were linked to the lives of individuals. Chapman has interpreted figurine fragmentation as widely distributed and frequent social practice, a routine part of everyday life in which the sacred and profane elements are present in different relation. Enchainment is one of such practices through which the living and their ancestors become connected by material means – signs of value which became figurine fragments (Chapman, 2000). Biehl considers figurine fragmentation as destruction, and destruction as well as creation he sees as communicative acts. Biehl (2006) considers this an intentional symbolic act.

Conclusions

Armenia, rich in archaeological sites and related funerary remains, has the potential to reconstruct the life of past societies. This study presents a survey of current evidence on human decapitation and decapitation of anthropomorphic figurines in the Early Bronze Age from the Armenian highland as well as a reinterpretation of a specific context. We infer that some individuals at the Landjik site were victims of ritualized decapitations. The results of the analyses conducted in this work show that the co-occurrence of decapitation of anthropomorphic figurines and human decapitation may not be random.

Acknowledgements

I would like to thank Gerhard Meisenberg for his helpful comments and editing. I thank the anonymous referees for their insightful and helpful comments.

References

Alekseev, V.P. & Debets, G.F. (1964). *Kraniometriya (metodika antropologicheskikh issledovaniy)* [Craniometry (methods of anthropological research)]. Moscow: Science.

Allen, C.J. (2002). *The Hold Life Has: Coca and Cultural Identity in an Andean Community*, 2nd edition. Washington, DC: Smithsonian Institution Press.

Ames, K.M. & Maschner, H.D.G. (1999). *The Peoples of the Northwest Coast: Their Archaeology and Prehistory.* London: Thames & Hudson.

Anderson, T. (2001). Two decapitations from Roman Towcester. *International Journal of Osteoarchaeology* 11: 400-405

Andrushko, V.A. & Torres, E.C. (2011). Skeletal evidence for Inca warfare from the Cuzco region of Peru. *American Journal of Physical Anthropology* 146: 361-372.

Anthony, D.W. (2010). *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton NJ: Princeton University Press

Ardagna, Y., Richier, A., Vernet, G., Dutour, O. (2005). A case of beheading dating from the Celtic Period (La Tène B. Sarliève-Grande Halle, France). *International Journal of Osteoarchaeology* 15: 73-76.

Aufderheide, A.C. & Rodriguez-Martin, C. (1998).*The Cambridge Encyclopedia of Human Paleopathology.* New York: Cambridge University Press.

Avetisyan, P.S. (2015). *Armenian Highland during the 24th-9thCenturies BC (The Dynamics of Socio-Cultural Transformations, According to Archaeological Data)*. Research report: On the degree of Doctor of Historical Sciences, Yerevan.

Becker, L. (1988). Human being: The boundaries of the concept. In: M.F. Goodman (ed.), *What Is a Person?* pp. 145-167. Clifton, NJ: Humana Press.

Bell, C. (1992). *Ritual Theory, Ritual Practice.* Oxford: Oxford University Press.

Benoît, F. (1975). The Celtic oppidum of Entremont, Provence. In: R. Bruce-Mitford (ed.), *Recent Archaeological Excavations in Europe*, pp. 227-259. London: Routledge & Kegan Paul.

Berryman, H.E. & Haun, S.J. (1996). Applying forensic techniques to interpret cranial fracture patterns in an archaeological specimen. *International Journal of Osteoarchaeology* 6: 2-9.

Biehl, P.F. (1996). Symbolic communication systems: Symbols on anthropomorphic figurines of the Neolithic and Chalcolithic from Southeastern Europe. *Journal of European Archaeology: Journal of the European Association of Archaeologists* 4: 153-176.

Biehl, P.F. (2006). Figurines in action: Methods and theories in figurine research. In: R. Layton, S. Shennan & P. Stone (eds.), *A Future for Archaeology*, pp. 199-216. London: Routledge Cavendish.

Bienert, H.D. (1991). Skull cult in the prehistoric Near East. *Journal of Prehistoric Religion* 5: 9-23.

Bloch, M. (1992). *Prey into Hunter: The Politics of Religious Experience.* Cambridge: Cambridge University Press.

Bonogofsky, M. (2001). *An Osteo-Archaeological Examination of the Ancestor Cult during the Pre-Pottery Neolithic B Period in the Levant*. Ph.D. dissertation, University of California, Berkeley. Ann Arbor University Microfilms International.

Bonogofsky, M. (2003). Neolithic plastered skulls and railroading epistemologies. *Bulletin of the American Schools of Oriental Research* 331: 1-10.

Bonogofsky, M. (2004). Including women and children: Neolithic modeled skulls from Jordan, Israel, Syria and Turkey. *Near Eastern Archaeology* 67(2): 118-119.

Borodovsky, A. & Tabarev, A. (2005). Scalping in North America and western Siberia: The archaeological evidence. *Archaeology, Ethnology & Anthropology of Eurasia* 1(21): 87-96.

Boylston, A., Knüsel, C.J., Roberts, C.A. & Dawson, M. (2000). Investigation of a Romano-British rural ritual in Bedford, England. *Journal of Archaeological Science* 27: 241-254.

Brickley, M. & Smith, M. (2006). Culturally determined patterns of violence: Biological anthropological investigations at a historic urban cemetery. *American Anthropologist* 1: 163-177.

Brothwell, D.R. (1981). *Digging Up Bones*, 3rd edition. Ithaca: Cornell University Press.

Buckberry, J.L. & Hadley, D.M. (2007). An Anglo-Saxon execution cemetery at Walkington Wold, Yorkshire. *Oxford Journal of Archaeology* 26: 309-329.

Burton, R. (1864). Notes on scalping. *Anthropological Review* 2: 49-52.

Bynum, C.W. & Gerson, P. (1997). Body part reliquaries and body parts in the Middle Ages. *Gesta* 36(1): 3-7.

Carmichael, P.H. (1994). The life from death continuum in Nasca imagery. *Andean Past* 4: 81-90.

Chapman, J. (2000). *Fragmentation in Archaeology: People, Places and Broken Objects in the Prehistory of South Eastern Europe.* London/New York: Routledge

Conklin, B.A. (2001). *Consuming Grief: Compassionate Cannibalism in an Amazonian Society*. Austin: University of Texas Press.

Connerton, P. (1989). *How Societies Remember*. Cambridge: Cambridge University Press.

De Laguna, F. (1933). Mummified heads from Alaska. *American Anthropologist* 35(4): 742-744.

Duara, T. (1988). Superscribing symbols: The myth of Guandi, Chinese god of war. *Journal of Asian Studies* 47: 778-795.

Duncan, W.N. (2005). Understanding veneration and violation in the archaeological record. In: G.F.M. Rakita, J.E. Buikstra, L.A. Beck & S.R. Williams (eds.), *Interacting with the Dead: Perspectives on Mortuary Archaeology for the New Millenium*, pp. 207-227. Gainesville: University Press of Florida.

Duncan, W.N. & Schwarz, K.R. (2015). A Postclassic Maya mass grave from Zacpetén, Guatemala. *Journal of Field Archaeology* 40(2): 143-165.

Eeckhout, P. & Owens, L.S. (2008). Human sacrifice at Pachacamac. *Latin American Antiquity* 19(4): 375-398.

Eliade, M. (1964). *Shamanism: Archaic Techniques of Ecstasy*, translated by W.R. Trask. London: Arkana.

Galloway, A. (1999). *Broken Bones: Anthropological Analysis of Blunt Force Trauma*. Springfield, IL: Charles C. Tomas.

Green, M. (1998). Humans as ritual victims in the later prehistory of Western Europe. *Oxford Journal of Archaeology* 17: 169-189.

Hamilton, L.A. (2005). *Cut Marks as Evidence of Precolumbian Human Sacrifice and Postmortem Bone Modification on the North Coast of Peru*. Unpublished Ph.D. dissertation, Department of Anthropology, Tulane University, New Orleans.

Harner, M.J. (1962). Jivaro souls. *American Anthropologist* 64: 258-272.

Hartnady, P. & Rose, J.C. (1991). Abnormal tooth-loss patterns among Archaic-period inhabitants of the lower Pecos region, Texas. In: M.A. Kelley & C.S. Larsen (eds.), *Advances in Dental Anthropology*, pp. 267-278. New York: Wiley Liss.

Hill, E. (2006). Moche skulls in cross-cultural perspective. In: M. Bonogofsky (ed.), *Skull Collection, Modification and Decoration*, pp. 91-100. Oxford: Archaeopress.

Hinard, F. (1985). *Les proscriptions de la Rome républicaine*. Rome: École Française de Rome (Publications de l'École française de Rome, 83).

Hughes, D. (1991). *Human Sacrifice in Ancient Greece*. London: Routledge.

Jasin, H.E. (2005). Mechanisms of tissue damage in rheumatoid arthritis. In: W.J. Koopman & L.W. Moreland (eds.), *Arthritis and Allied Conditions. A Textbook of Rheumatology*, pp. 1141-1164. Philadelphia: Lippincott Williams & Wilkins.

Judd, M. (2004). Trauma in the city of Kerma: Ancient versus modern injury patterns. *International Journal of Osteoarchaeology* 14: 34-51.

Judd, M. (2006). Continuity of interpersonal violence between Nubian communities. *American Journal of Physical Anthropology* 131: 324-333.

Kennedy, G.E. (1986). The relationship between auditory exostoses and cold water: A latitudinal analysis. *American Journal of Physical Anthropology* 71: 401-415.

Kenyon, K. (1957). *Digging Up Jericho*. New York: Praeger.

Khudaverdyan, A. (1996). Antropologiya drevnego Landjikskogo naseleniya [Anthropology of the ancient population from Landjik]. In: A. Khalantaryan & S.

Harutyunyan (eds.), *10 Archaeological Sessions (1993-1995),* pp. 33-34. Abstract. Yerevan.

Khudaverdyan, A. (2011). Migrations in the Eurasian steppes in the light of paleoanthropological data. *Mankind Quarterly* 51: 387-463.

Khudaverdyan, A.Yu. (2012). Nonmetric cranial variation in human skeletal remains from Armenian Highland: Microevolutionary relations and intergroup analysis. *European Journal of Anatomy (Spanish)* 16 (2): 134-149.

Khudaverdyan, A.Yu. (2013). A dental nonmetric analysis of Bronze Age population from Armenian Plateau. *Anthropological Review* 76(1): 63-82.

Khudaverdyan, A.Yu. (2014a). Ritual and ceremonial dismembering bones in burials in Bronze and Iron centuries from Armenian Plateau. *Journal of Antropologija* 2(14): 163-195.

Khudaverdyan, A.Yu. (2014b). Decapitations in Late Bronze Age and Iron Age sites from Sevan region (Armenia). *Journal of Siberian Federal University. Humanities & Social Sciences* 7(9): 1555-1566.

Khudaverdyan, A.Yu. (2015a). Palaeopathology of human remains of the 1st century BC–3rd century AD from Armenia (Beniamin, Shirakavan I). *Anthropological Review* 78: 213-228.

Khudaverdyan, A.Yu. (2015b). Decapitation in the Dvina (Armenia): An analysis of the skulls of Middle Ages. *Journal of Paleopathology* (Italy) 25(1-3): 25-34.

Khudaverdyan, A.Yu. (2016). Late Bronze Age and Iron Age crania from Armenia: A paleoecological study. *Archaeology, Ethnology & Anthropology of Eurasia* 2(44): 129-136.

Khudaverdyan, A.Yu., Petrosyan, L.A., Khachatryan, H.H. & Yeganyan, L.G. (2015). Antropologicheskie materiali iz pogrebenii kura-araksskoj kulturi s territorii Shirakskoi ravnini. [Anthropological materials from burials of the Kura-Araks culture from territory of the Shirak plain]. *Centr armenovedcheskikh issledovanii Shiraka* [Armenology Research Center. Scientific Works] XVIII: 5-22.

Khudaverdyan, A., Devedjian, S., Vardanyan, Sh. & Yengibaryan, A. (2016). The extraordinary burial of Lori Berd memorial: Possible interpretations. *Aramazd – Armenian Journal of Near Eastern Studies* 2, in press.

Knight, B. (1991). *Forensic Pathology*. New York: Oxford University Press.

Larsen, C.S. (1997). *Bioarchaeology: Interpreting Behavior from the Human Skeleton*. Cambridge: Cambridge University Press.

Last, J. (1998). A design for life: Interpreting the art of Catalhöyük. *Journal of Material Culture* 3: 355-378.

Lintott, A.W. (1968). *Violence in Republican Rome*. Oxford: Oxford University Press.

Lovell, N.C. (1997). Trauma analysis in paleopathology. *American Journal of Physical Anthropology* 104(25): 139-170.

Malone, C. (1998). God or goddess: The temple of Malta. In: L. Goodison & C. Morris (eds.), *Ancient Goddesses: The Myths and Evidence*, pp. 148-163. London: British Museum Press.

Manchester, K. (1983). *The Archaeology of Disease*. Bradford.

Manzi, G., Sperduti, A. & Passarello, P. (1991). Behaviour-induced auditory exostoses in Imperial Roman society: Evidence from coeval urban and rural communities near Rome. *American Journal of Physical Anthropology* 85: 253-260.

Martin, D.L. & Akins, N.J. (2001). Unequal treatment in life as in death: Trauma and mortuary behavior at La Plata (AD 1000–1300). In: D.R. Mitchell & J.L. Brunson-Hadley (eds.), *Ancient Burial Practices in the American Southwest*, pp. 223-248. Albuquerque: University of New Mexico Press.

Maschner, H.D.G. (1999). Prologue to the prehistory of the Lower Alaska Peninsula. *Arctic Anthropology* 36(1-2): 84-102.

Matić, U. (2009). Power over the body in a hybrid reality: Anthropomorphic figurines of Bubanj-Salcuţakrivodol complex on the Central Balkans. In: 15th Annual Meeting of the European Association of Archaeologists in Riva del Garda (Trento, Italy), 15th-20th September 2009, Abstracts p. 202.

McKinley, J.I. (1993). A decapitation from the Romano-British cemetery at Baldock, Hertfordshire. *International Journal of Osteoarchaeology* 3: 41-44.

Meindl, R.S., Lovejoy, C.O., Mensforth, R.P. & Carlos, L.D. (1985). Accuracy and direction of error in the sexing of the skeleton: Implications for paleodemography. *American Journal of Physical Anthropology* 68: 79-85.

Mellaart, J. (1967). *Catal Huyuk: A Neolithic Town in Anatolia*. London: Thames & Hudson.

Millaire, J.-F. (2015). *The Sacred Character of Ruins on the Peruvian North Coast*. Manuscript on file, University of Western Ontario, London, Canada.

Moser, C.L. (1973). *Human Decapitation in Ancient Mesoamerica. Studies in Pre-Columbian Art and Archaeology 11*. Washington DC: Dumbarton Oaks.

Movsesyan, A.A., Mamonova, N.N. & Richkov, Yu.G. (1975). The program and method of research of anomalies of a skull. *Anthropology Questions* 51: 127-150.

Murphy, R.F. (1959). Intergroup hostility and social cohesion. *American Anthropologist* 59(6): 1018-1035.

Nagaoka, T. & Abe, M. (2007). Human skeletal remains from the Osaka castle site in Japan: Metrics and weapon injuries. *Anthropological Science* 115: 163-168.

Oakdale, S. (2001). History and forgetting in an indigenous Amazonian community. *Ethnohistory* 48(3): 381-401.

Okumura, M.M.M. & Eggers, S. (2008). Natural and cultural formation processes on the archaeological record: A case study regarding skeletal remains from a Brazilian shell mound. In: A.R. Suárez & M.N. Vásquez (eds.), *Archeology Research Trends*, Chapter 1, pp. 1-39. New York: Nova Science.

Onians, R. (1973). *The Origins of European Thought*. New York: Arno Press.

Ortner, D.J. (2003). *Identification of Pathological Conditions in Human Skeletal Remains*, 2nd ed. Academic Press.

Paul, A. (2001). Bodiless heads in Paracas necropolis textile iconography. *Andean Past* 6: 69-94.

Pearson, P.M. (1999). *The Archaeology of Death and Burial*. College Station: Texas A&M University Press.

Peters, A.H. (1991). Ecology and society in embroidered images from the Paracas necropolis. In: A. Paul (ed.), *Paracas Art and Architecture: Object and Context in South Coastal Peru*, pp. 240-313. Iowa City: University of Iowa Press.

Petres, É.F. (1972). On Celtic animal and human sacrifices. *Acta Archaeologica Academiae Scientiarum Hungaricae* 24(4): 365-383.

Petrosyan, L.A. (1996). Pekhumner Landjikum [Excavations in Landjik]. In: A. Khalantaryan & S. Harutyunyan (eds.), *10 Archaeological Session (1993-1995),* pp. 60-61. Abstract. Yerevan.

Proulx, D.A. (2001). Ritual uses of trophy heads in ancient Nasca society. In: E.P. Benson & A.G. Cook (eds.), *Ritual Sacrifice in Ancient Peru*, pp. 119-136. Austin: University of Texas Press.

Riches, D. (1986). The phenomenon of violence. In: D. Riches (ed.), *The Anthropology of Violence*, pp 1-27. Oxford: Basil Blackwell.

Richlin, A. (1999). Cicero's head. In: J.I. Porter (ed.), *Constructions of the Classical Body*, pp. 190-211. Ann Arbor: University of Michigan Press.

Sauer, N.J. (1998). The timing of injuries and manner of death: Distinguishing among antemortem, perimortem and postmortem trauma. In: K. Reichs (ed.), *Forensic Osteology*, pp. 321-332. Springfield, IL: Charles C. Thomas.

Schmandt-Besserat, D., Griffin, P.S., Grissom, C.A., Rollefson, G.O. & Rose, J.C. (2010). *From Behind the Mask: Plastered Skulls from Ain Ghazal*. Available at http://menic.utexas.edu/ghazal/ChapV/skull/ .

Seeman, M.F. (1988). Ohio Hopewell trophy-skull artifacts as evidence for competition in Middle Woodland societies circa 50 B.C.–A.D. 350. *American Antiquity* 53: 565-577.

Smith, M.O. (1997). Osteological indications of warfare in the Archaic Period of the Western Tennessee Valley. In: D.L. Martin & D.W. Fayer (eds.), *Troubled Times: Violence and Warfare in the Past*, pp. 241-266. Amsterdam: Gordon & Breach.

Stokstad, M. (2002). *Art History,* vol. 1. Upper Saddle River, NJ: Prentice-Hall.

Talalay, L.E. (2004). Heady business: Skulls, heads, and decapitation in Neolithic Anatolia and Greece. *Journal of Mediterranean Archaeology* 17(2): 139-163.

Taylor, A.C. (1993). Remembering to forget: Identity, mourning and memory among the Jivaro. *Man* n.s. 28: 653–678.

Thorpe, I.J.N. (2003). Anthropology, archaeology, and the origin of warfare. *World Archaeology: The Social Commemoration of Warfare* 35(1): 145-165.

Todd, M. (1987). *The Northern Barbarians: 100 BC–AD 300*, revised edition. Oxford: Basil Blackwell.

Ucko, P.J. (1962). The interpretation of prehistoric anthropomorphic figurines. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 92(1): 38-54.

Ucko, P.J. (1968). *Anthropomorphic Figurines of Predynastic Egypt and Neolithic Crete with Comparative Material from the Prehistoric Near East and Mainland Greece*. London: A. Szmidla.

Urcid, J. (1994). Cannibalism and curated skulls: Bone ritualism on Kodiak Island. In: T.L. Bray & T.W. Killion (eds.), *Reckoning with the Dead: The Larsen Bay Repatriation and the Smithsonian Institution*, pp. 101-121. Washington DC: Smithsonian Institution Press.

Van Dyke, R.M. & Alcock, S.E. (2003). *Archaeologies of Memory*. Oxford: Blackwell.

Verdery, K. (1999). *The Political Lives of Dead Bodies: Reburial and Postsocialist Change*. New York: Columbia University Press.

Waldron, T. (1996). Legalized trauma. *International Journal of Osteoarchaeology* 6: 114-118.

Walker, P.L. (2001). A bioarchaeological perspective on the history of violence. *Annual Review of Anthropology* 30: 573-596.

Williams, H. (2001). Death, memory and time: A consideration of the mortuary practices at Sutton Hoo. In: C. Humphrey & W.M. Ormrod (eds.), *Time in the Medieval World*, pp. 35-71. New York: York Medieval Press.

Williams, H. (2003a). *Archaeologies of Remembrance: Death and Memory in Past Societies*. New York: Kluwer Academic/Plenum.

Williams, H. (2003b). Remembering and forgetting the medieval dead: Exploring death, memory and material culture in monastic archaeology. In: H. Williams (ed.), *Archaeologies of Remembrance: Death and Memory in Past Societies*, pp. 227-254. New York: Kluwer Academic/Plenum.

Williams, S.R., Forgey, K. & Klarich, E. (2001). An osteological study of Nasca trophy heads collected by A. L. Kroeber during the Marshall Field expeditions to Peru. *Fieldiana: Anthropology* n.s. 33: 1-132.

Wilson, C.E. (1981). Burials within settlements in Southern Britain during the pre-Roman Iron Age. *Bulletin of the Institute of Archaeology* 18: 127-170.

Zubov, A.A. (1968). *Odontology: A Method of Anthropological Research*. Moscow: Science.

MANKIND QUARTERLY 2017 57:4 519-541

Differences in Intelligence and Socio-Economic Outcomes across the Twenty Seven States of Brazil

Richard Lynn
University of Ulster, UK

Mayra Antonelli-Ponti,* Rodrigo Focosi Mazzei,
José Aparecido Da Silva
Psychobiology Unit, University of São Paulo at Ribeirão Preto, Brazil

Gerhard Meisenberg
Ross University School of Medicine, Dominica

*Corresponding author: antonelli.may@gmail.com

In a number of countries, earlier studies have reported significant associations between regional differences in intelligence within countries and economic and social phenomena. Using scores on the Program of International Student Assessment (PISA) tests as indicator of intelligence, we find statistically significant correlations for the 27 states of Brazil between intelligence and nine indicators of socio-economic development. Spatial analysis indicates that relationships are present both at the level of differences between adjacent states and over long-distance clines. Most of the relationships observed after initial analysis persisted after controlling for spatial autocorrelation. Among the socio-economic variables, those that describe the standard of living of the less affluent sections of the population tend to correlate most with PISA scores.
Key words: Intelligence; PISA; Brazil; Income; Health; Fertility; Spatial analysis

The importance of intelligence for people's lives has been shown in numerous studies that found high correlations of intelligence, measured by IQ tests, with many important outcomes. Measured intelligence is considered the best predictor of school performance (Deary et al., 2007; Laidra, Pullmann & Allik, 2007; Rindermann & Neubauer, 2004), being more potent than other plausible predictors (Poropat, 2009). Importantly, cognitive abilities measured early in life are great predictors of future events such as health and longevity (Gottfredson & Deary, 2004), level of schooling, job placement, and income (Sorjonen et al., 2015). The implication of this is that cognitive ability is not only an outcome of circumstances in a person's life, but to some extent also a cause.

This conclusion is plausible theoretically, in addition to being compatible with the observed relationships. Problem solving ability is part of intelligence (Carroll, 1993; Snyderman & Rothman, 1988), and it impacts all aspects of life: accident prevention, study strategies, struggling with information technology, navigating the job market, and managing one's personal life. Problem solving ability is an essential part of competence in virtually every occupation. In this way, people tend to cluster intellectually and create their personal environments according to their intellectual level and personal preferences (Da Silva, Ribeiro-Filho & Santos, 2012).

The importance of intelligence, as well as other personal qualities not examined in our research, is not confined to the personal lives of individuals. These traits also shape the communities that people form and the worldviews, lifeways and forms of governance they adopt. Intelligence is postulated to be a major driving force of economic and cultural development (Meisenberg, 2014; Rindermann, Sailer & Thompson, 2009). In support of this hypothesis, many studies found that regional differences in intelligence within countries are robustly associated with economic and social phenomena (Table 1, see also Lynn & Vanhanen, 2012). Davenport and Remmers (1950) were the first to report relationships of state IQ with per capita income, lynchings, and several other outcomes in the United States. The positive association with income was confirmed and followed up by Kanazawa (2006) and McDaniel (2006) more than half a century later.

Further studies reporting associations of intelligence with per capita income and related economic measures at the level of sub-national administrative units have been conducted in other countries (Table 1). They include 13 regions of the British Isles (Lynn, 1979), 90 regions of France (Lynn, 1980; Montmollin, 1958), 12 regions of Italy (Lynn, 2010), 19 regions of Italy (Piffer & Lynn, 2014; Templer, 2012), 18 regions of Spain (Lynn, 2012), 16 regions of Germany (Roivainen, 2012), 47 prefectures of Japan (Kura, 2013), 12 regions of Turkey (Lynn, Sakar & Cheng, 2015), 31 regions of the People's Republic of China (Lynn & Cheng, 2013), and 12 regions of the United Kingdom (Carl, 2016). Many of these studies also reported significant correlations between regional differences in intelligence and a variety of social phenomena including health, fertility and crime.

Table 1.*Summary of earlier studies reporting correlations between intelligence of regions and socioeconomic indicators. Correlations are for measures of per capita GDP, average income, or related measures.*

Country	Outcomes reported	Author, date	*r* with GDP/income
USA	12 economic and social variables	Davenport & Remmers,1950	0.32
USA	Gross domestic product (GDP)/capita	Kanazawa, 2006	0.57
USA	GDP, health, crime, govt. effectiveness	McDaniel, 2006	0.59
Britain	GDP, unemployment, infant mortality, crime	Lynn, 1979	0.73
France	Earnings, unemployment, infant mortality, migration	Lynn, 1980	0.61
Italy	Income, education, infant mortality	Lynn, 2010	0.94
Italy	GDP, murder, anthropometry	Templer, 2012	0.98
Spain	Income, employment, literacy, life expectancy	Lynn, 2012	0.42
Germany	Education, church membership	Roivainen, 2012	0.79 (1992) 0.27 (1998)
China	GDP, ethnicity, education	Lynn & Cheng, 2013	0.42
Japan	Income, homicide, divorce, fertility, suicide	Kura, 2013	0.51
Turkey	Income, education, fertility, infant mortality	Lynn, Sakar & Cheng, 2015	0.81
Britain	Unemployment, poverty, patent applications	Carl, 2016	0.42

In our study we use the scores of PISA tests as measures of intelligence, forming the average of the Reading, Mathematics and Science sections. PISA scores are adequate measures of intelligence. Scientific competence, language skills and mathematics, along with reasoning and problem solving, are components of general intelligence in Carroll's (1993, p. 524) taxonomy. Carroll identifies math ability as "quantitative reasoning". Among other components (1993, pp. 598-599) he lists reading comprehension, defined in the PISA studies as the capacity to understand, use and reflect on written texts, and science understanding identified as "general science information". Furthermore, it has been inferred that the same genetic polymorphisms determine cognitive ability measured by educational tests and intelligence tests (Bartels et al., 2002; Plomin, Kovas & Haworth, 2007).

Furthermore, the data deal with populations, not individuals. Therefore, we will always work with averages. The aim is to relate the average intelligence of state populations to indicators that represent important real-life phenomena. The aim is to chart the relationships between the cognitive traits of local populations and the environments created by them (Jones, 2016). The average, therefore, does not represent any individual, and the observed relationships are population-level phenomena that may or may not hold true at the individual level. Our hypothesis is that intelligence has a positive impact on desirable conditions, and a negative impact on undesirable conditions. More specifically, intelligence of Brazilian states is predicted to have a moderate or strong positive correlation with higher education, income, life expectancy and water availability, and a moderate or strong negative correlation with infant mortality, fertility, poverty and violence.

In Brazil, the economists Curi and Menezes (2014) investigated the association of income with mathematical abilities measured in an educational testing program (SAEB - Assessment of the Basic Education System) at the level of Brazilian states. Fuerst and Kirkegaard (2016) investigated ancestry factors of Latin American countries, including Brazil; Kirkegaard (2015) also investigated PISA scores and their associations, but focusing on socioeconomic development. Our study adds another piece to the puzzle of intelligence and its correlates in the provinces of a large country. It is a study with a view of the real-world implications of the psychological construct of intelligence. Considering the geographic scale of Brazil, where each state can have the size of a European country, the subject of the study is of more than purely local interest. Its results can point to the social importance of developing people’s cognitive abilities, and perhaps provide indications of how this aim can be achieved.

Method

The measures of intelligence for the 27 states (26 federal states and the federal district) of Brazil (Fig. 1) were calculated from the PISA (Programme for International Student Assessment) scores of 15-year-old school students in Math, Science and Reading for the years 2009, 2012 and 2015 with results for 61,862 Brazilian school students (OECD, 2016). The values were obtained as follows:

- The plausible values (five in 2009 and 2012 and ten in 2015) of each area (mathematics, reading and science) were obtained by state and for each of the three years.
- The average of plausible values was calculated for each state and year.

- The average of the total scores was calculated by state and year.
- The average of the three years was calculated for each state with weighting by sample size, which was different in different years.

The PISA data have been used as measures of intelligence in a number of studies, e.g. for the regions of Italy (Lynn, 2010; Templer, 2012) and of Turkey (Lynn, Sakar & Cheng, 2015). These tests are regarded as valid measure of intelligence with which they are highly correlated (Kaufman et al., 2012). It has been shown that PISA tests and intelligence tests are very highly correlated across countries, e.g. at r=.89 for 63 countries reported by Rindermann (2007) and at r=.91 for 82 countries reported by Meisenberg and Lynn (2011).

The economic and social data for 2010 were collected from a Brazilian national database (Atlas of Human Development in Brazil, 2013). The violence data were obtained for 2014 and 2015 from State Departments of Public Safety and/or Social Defense; National Statistical System for Public Security (SINESP); Brazilian Institute of Geography and Statistics (IBGE); and Brazilian Forum on Public Security (Folha de São Paulo, 2015; O Estado de São Paulo, 2016).

1. *Higher education*: Percentage with a university degree.
2. *Income*: Per capita household income in Brazilian currency reais.
3. *Infant mortality*: Deaths of infants below 1 year per 1,000 live births.
4. *Water availability*: Percentage of the population living in houses with piped water.
5. *Life expectancy*: Average number of years that individuals born in the same year can expect to live, for men and women.
6. *Fertility*: Total Fertility Rate (TFR) defined as the average number of children a woman would have in her life at current age-specific fertility rates.
7. *Poverty*: Percentage of individuals with an income of 70 reais or less per month.
8. *Violence 2014*: Intentional lethal violent crimes, including homicide, robbery, bodily injury followed by death and victims of police actions in 2014 (Folha de São Paulo, 2015).
9. *Violence 2015*: Intentional lethal violent crimes, including homicide, robbery, bodily injury followed by death and victims of police actions in 2015 (O Estado de São Paulo, 2016).

Descriptive statistics for PISA scores and the economic and social variables are given in Appendix 1.

For the analysis of the data, we chose correlation, spatial analysis and principal components analysis. Statistical data analysis was done with SPSS-23 statistics software.

Figure 1. *Map of the 26 Brazilian states and the Federal District of Brasilia.*

Results

Correlations

The Pearson product moment correlations between the variables were examined first. The results are given in Table 2. The correlations above the diagonal were calculated without weighting, and those below the diagonal are weighted for the square root of population size, converted to a weight factor with average value of 1 in order to preserve degrees of freedom for statistical analysis. Weighting was done because small states are more likely than larger ones to be affected by idiosyncratic conditions and to have atypical relationships between the variables. For example, the small Federal District might be atypical on some measures because of its high concentration of well-paid government officials. Table 2 shows that correlations tend to be somewhat higher with weighting than without, as is expected if the observed relationships are true and if the signal-to-noise ratio is higher in the bigger states.

Table 2. *Correlation matrix for PISA scores and economic and social variables. Unweighted correlations are above the diagonal, and correlations weighted by square root of population size are below the diagonal. N = 27 states; *, p<.05; **, p<.01; ***, p<.001, two-tailed t tests.*

	PISA score	Higher educ.	Income	Infant mort.	Water available	Life exp.	Fertility	Poverty	Violence 2014	Violence 2015
PISA score	1	.721***	.791***	-.815***	.765***	.869***	-.678***	-.854***	-.475*	-.499**
Higher education	.739***	1	.964 ***	-.658***	.613***	.760***	-.480*	-.724***	-.426*	-.497**
Income	.786***	.972***	1	-.720***	.698***	.803***	-.579***	-.811***	-.414*	-.488**
Infant mortality	-.875***	-.720***	-.782***	1	-.693***	-.873***	.522**	.838***	.506**	.454*
Water available	.813***	.712***	.765***	-.789***	1	.734***	-.887***	-.891***	-.172	-.259
Life expectancy	.878***	.764***	.793***	-.877***	.774***	1	-.564***	-.845***	-.493**	-.514**
Fertility	-.715***	-.620***	-.674***	.652***	-.901***	-.614***	1	.706***	.075	.165
Poverty	-.874***	-.793***	-.856***	.897***	-.932***	-.855***	.770***	1	.313	.370
Violence 2014	-.614***	-.540**	-.550**	.576**	-.387*	.576**	.332	.499**	1	.930***
Violence 2015	-.636***	-.621***	-.624***	.560**	-.472*	.604***	.426*	.559**	.947***	1

Most of the correlations of PISA scores with indicators of economic prosperity (income, poverty, water availability) and health (infant mortality, life expectancy) are near .80 or above .80, and correlations with higher education and fertility are not much lower. Also the socio-economic variables are highly correlated among each other as expected, to an extent similar to their correlations with PISA scores. Correlations of the two measures of violence are lower both with PISA scores and with the other socio-economic indicators.

Spatial patterning

The statistical significance values in Table 2 must not be taken at face value because of spatial patterning of the variables. The usual observation in geographic data is that neighboring units are similar on many dimensions. The greater the distance, the greater are the differences. The implication for data analysis is that the data points are not independent of each other and that traditional significance testing, which assumes independence of data points, is not applicable (Eff, 2004; Hassall & Sherratt, 2011). This is especially the case when features such as average PISA score, income or life expectancy form a geographic gradient, or cline, over an extended territory. In Brazil, the southern states in general tend to have higher levels of socio-economic development than the northern states. For example, the correlation with latitude is -.787 for PISA score, -.656 for income, and -.709 for life expectancy. Any variable that shows clinal variation will produce high correlations with any other variable that forms a cline in a similar direction. Indeed, two perfect clines, lying at a 45º angle, will produce a correlation of r=.71. In other words, two random clines have a 50% probability of being correlated at r>.71.

To explore the extent to which the results are due to spatial gradients, we calculated for each variable the differences between each of the 351 pairs of states, and related them to the distance between the states (proxied by their capital cities). The positive signs in the first data column of Table 3 show that as expected, for each of the variables, states that are more distant are more different from each other. This clinal distribution is most evident for fertility, availability of piped water and poverty, but minimal for percentage with higher education degrees. The distance effect is intermediate for PISA scores.

A more commonly used measure of spatial patterning is Moran's *I*, which is defined by the following formula:

$$\text{Moran's } I = (N / \textstyle\sum_{i,j} wij) \times (\textstyle\sum_{ij}(wij \times zXi \times zXj) \,/\, \textstyle\sum_i zXi^{\,2})$$

N is the number of state pairs (351), and w_{ij} is the spatial weight. Distances ranged from 104 km (Paraiba – Pernambuco) to 3788 km (Roraima – Rio Grande do Sul), and we calculated the spatial weight as inverse distance between the state capitals: distance subtracted from 3892. *zXi* and *zXj* are the standardized values of the variable in state i and state j. Numerically, Moran's *I* can vary between -1 and +1. The values of Moran's *I* for each of the 10 variables are listed in the third data column of Table 3. Across the 10 variables, the correlation of Moran's *I* with the simple distance-difference correlation in column 1 is .926 (p<.001). Thus these two measures of spatial dependence are nearly equivalent.

Another question, and one that is more relevant for statistical analysis, is whether the relationship between two focal variables is consistent across locations. To this end we performed simple regressions across the 27 provinces with PISA score as independent variable predicting each of the other 9 variables in Table 3. The differences between the residuals were then tabulated for the 351 pairs of states and correlated with distance. Column 2 in Table 3 shows that results are quite different for different variables, with income and higher education showing significant negative correlations. Across the 9 variables, the distance-dependence of residuals correlated .706 (p = .034) with the distance-dependence of the scores and .697 (p = .037) with Moran's *I*. The result suggests that in Brazil, some of those socio-economic indicators that show more pronounced long-distance clines (fertility, water availability) also tend to have greater distance-dependent differences in their non-cognitive determinants.

Table 3. *Spatial patterning of variables in the Brazilian states. N = 27 states; *, p<.05; **, p<.01; ***, p<.001, two-tailed t tests.*

Measure	Correlation of distance with		Moran's *I*	
	score difference	residual difference	Scores	diff. scores
PISA score	0.260***		0.129	-0.049
Higher education	0.024	-0.152**	0.041	-0.060
Income	0.151**	-0.184***	0.080	-0.057
Infant mortality	0.245***	0.105	0.098	-.044
Water available	0.496***	0.187***	0.158	-0.025
Life expectancy	0.262***	0.032	0.116	-0.057
Fertility	0.503***	0.410***	0.156	-0.006
Poverty	0.426***	-0.012	0.160	-0.024
Violence 2014	0.145**	0.113*	0.114	0.032
Violence 2015	0.150**	0.063	0.095	-0.018

The positive values for Moran's *I* emphasize the need to control for the non-independence of data points in order to obtain valid statistical significance estimates. The procedure followed was to form, for each variable, the difference score between the state's value and the average of the values of those states with which the state shares a boundary (Appendix 2). For example, the difference score for the fertility rate in Espirito Santo is the difference between fertility in Espirito Santo and the average of the fertility rates in Bahia, Minas Gerais and Rio de Janeiro (see Fig. 1). Because only neighboring states are compared, spatial proximity is no longer an important determinant of similarity between the states, and the correlations are no longer confounded – and inflated – by variations in spatial proximity. This method is similar to the K nearest neighbor method (Fortin, Dale & VerHoef, 2002), but avoids the need to somewhat arbitrarily decide about the number of nearest neighbors with which each state is compared. It determines whether, for example, states in which the average PISA score is higher than in the surrounding states also have higher life expectancy than the surrounding states.The last column in Table 3 shows that when Moran's *I* is calculated based on the difference scores, the values are closer to zero for all variables except higher education. The average Moran's *I* across the 10 variables is .115 for the calculation based on actual scores, and -.031 for the difference scores.

Table 4 shows that, as expected, the correlations between the difference scores are lower than the raw correlations in Table 2. However, most of the relationships remain moderately strong and many remain statistically significant. We see, for example, that states in which the PISA score is higher than in the surrounding states tend to have more piped water, higher life expectancy, lower fertility, lower infant mortality and lower poverty rates than the surrounding states, and that these relationships are unlikely to be chance observations. Income and higher education still are highly correlated, as are, obviously, rates of violent crime in 2014 and 2015. Now, income, higher education and violent crime are only weakly related to PISA scores and socio-economic indicators, but high correlations are preserved between members of a "poverty cluster" that includes poverty, water availability, infant mortality and fertility. This time we do not observe conspicuously higher correlations with weighted than unweighted data, presumably because greater "typicality" of larger states is offset by greater distances from the population centers of neighboring states. In general, the results presented in Table 4 support the existence of true relationships among many of the variables.

Principal components analysis was performed with the unweighted difference scores of the socioeconomic variables in Table 4, but excluding PISA scores. It produced three principal components with eigenvectors above 1. Figure 2 shows the scree plot. Table 5 shows the loadings of the variables on these three varimax-rotated principal components. The first of these can be described as a "poverty factor", with highest loadings on water availability, poverty, fertility and infant mortality. The second is a "prosperity factor" defined by income and proportion with a university education; and the third is a "crime factor". Also shown in Table 4 are the correlations of PISA scores with these three principal components. PISA scores are inversely associated with poverty. Correlations with the other principal components are of smaller magnitude and do not reach statistical significance with our sample size of 27 states, but are in the hypothesized direction.

Table 4. *Correlation between the difference scores. Unweighted correlations are above the diagonal, and correlations weighted by square root of population size are below the diagonal. N = 27 states; *, p<.05; **, p<.01; ***, p<.001, two-tailed t tests.*

	PISA score	Higher educ.	Income	Infant mort.	Water available	Life exp.	Fertility	Poverty	Violence 2014	Violence 2015
PISA score	1	.281	.280	.516**	.319	.509**	-.424*	-.436**	-.273	-.059
Higher education	.224	1	.933***	-.300	.288	.448*	-.184	-.357	-.330	-.362
Income	.177	. 946***	1	-.330	.336	.436*	-.363	-.419*	-.162	-.188
Infant mortality	-.526**	-.306	-.338	1	-.703***	-.659***	.674***	.753***	.198	-.089
Water available	.369	.331	.353	-.688***	1	.474*	-.766***	-.844***	-.171	-.096
Life expectancy	.502**	.302	.244	-.571**	.410*	1	-.380	-.493**	-.047	.112
Fertility	-.375	-.301	-.434*	.683***	-.732***	-.236	1	.702***	.157	.014
Poverty	-.443*	-.405*	-.454*	.760***	-.876***	-.411*	.740***	1	.074	-.058
Violence 2014	-.326	-.316	-.196	.149	-.166	.036	.271	.116	1	.844***
Violence 2015	-.176	-.373	-.244	-.072	-.121	.156	.182	.027	.876***	1

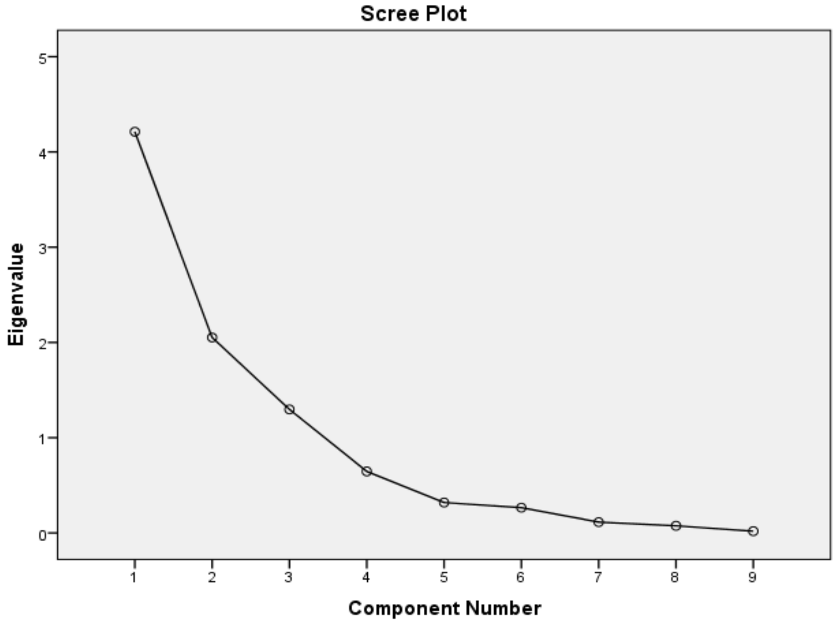

Figure 2. *Principal components analysis of unweighted difference scores, scree plot.*

Table 5. *Loadings of the difference scores on the first three principal components, **, p<.01.*

	Principal Component			Communalities
	1 "Poverty"	2 "Prosperity"	3 "Crime"	
Higher education	-.132	.944	-.259	.976
Income	-.222	.928	-.092	.919
Infant mortality	.868	-.204	-.015	.795
Water availability	-.905	.121	-.119	.847
Life expectancy	-.550	.510	.176	.594
Fertility	.863	-.064	.082	.756
Poverty	.879	-.235	-.034	.828
Violence 2014	.132	-.089	.938	.905
Violence 2015	-.076	-.137	.957	.941
Factor correlation with PISA score	-.528**	.291	-.231	

The causal paths between cognitive ability and the different outcomes are uncertain. One plausible hypothesis is that higher cognitive ability produces prosperity, indexed by income, and that all other outcomes are a direct consequence of higher income. In this case, income would beat PISA scores as the principal predictor of the other outcomes in multiple regression models. The test of this hypothesis is shown in Table 6. The β coefficients in the first two data columns show that most outcomes are independently predicted by both PISA scores and income, and that in most cases the coefficients are larger for PISA scores than for income. This indicates that either cognitive resources are more important than material resources for outcomes such as infant mortality and life expectancy, or that PISA scores are more accurate than income for measuring their respective constructs. The last two columns in Table 6 show the results when difference scores are used. Now the effect sizes and significance levels are generally reduced, but many of the PISA effects retain at least marginal statistical significance.

Table 6. *Results of regression models predicting diverse outcomes with PISA scores and income. Standardized β coefficients are shown. The first two data columns use actual scores, and the last two columns use difference scores for both the independent and dependent variables. Cases are weighted by square root of population size. N = 27 states; +, p<.10; *, p<.05; **, p<.01; ***, p<.001, two-tailed t tests.*

	Scores predicted by		Δ scores predicted by	
	PISA	Income	Δ PISA	Δ Income
Higher education	-.066	1.024***	.059	.936***
Infant mortality	-.681***	-.247	-.481**	-.253
Water availability	.552**	.331+	.316+	.297
Life expectancy	.668***	.268+	.474*	.161
Fertility	-.485*	-.293	-.308+	-.379*
Poverty	-.528**	-.441**	-.375*	-.389*
Violence 2014	-.476+	-.176	-.301	-.143
Violence 2015	-.380	-.325	-.137	-.220

Discussion

The main result of the present study is that cognitive ability, proxied by PISA scores of 15-year-olds, is related to a variety of socio-economic outcomes at the level of Brazilian states. Geographic distance is important, as is seen when the correlations in Table 2 are compared with those in Table 4. In general, we observe a pattern of attenuated but still substantial correlations for most relationships in Table 4. Correlations of PISA scores with infant mortality, life expectancy, poverty and fertility still are statistically significant, and the others are still in the same directions as in the original analysis. The results show that variation exists both between adjacent states and along long-distance clines, and that a control for spatial autocorrelation is therefore meaningful.

The results do not show that those correlations that became statistically non-significant in the spatial autocorrelation control should be dismissed. The spatial autocorrelation control that we used removes distance as a confounding factor by limiting the analysis to variance among neighboring states. The raw correlations reported in Table 2 are still relevant because they include the entire variance, including the long-distance component. Turning to specific results, we note six findings of interest:

1. The raw data show a high positive correlation between the intelligence of states and household income, with Pearson's *r* close to .79. Flores-Mendoza et al. (2012) analyzed the cognitive performance of Brazilian adults in four states: Amazônia, Bahia, Minas Gerais and São Paulo. They concluded that intelligence is not a good indicator for economic factors. Although our initial results reported in Table 2 seem to contradict this conclusion, the analysis of difference scores reported in Table 4 casts doubt on a key role of cognitive ability for income differences, at least between neighboring states. In Brazil, both cognitive ability and average income appear to vary mainly along correlated long-distance clines, rather than between neighboring states.

Associations between intelligence and income have been reported at the level of sub-national units for the United States, the British Isles, France, Italy, Spain, Germany, the People's Republic of China, and Turkey. However, none of these studies subjected its results to spatial analysis to clarify at what geographic scale these correlations arise. In the case of Brazil, the high correlation between PISA scores and income is easily explained considering that Brazil is an extremely unequal country. If we double the per capita GDP of the lowest-income state in 2010 (Maranhão - R$360.34), we do not even come close to the value of the state with the second highest income (São Paulo - R$1,084.46) in the same year. This does not take into account the Federal District with the capital Brasilia, which has the highest income (R$1715.11). Also average PISA scores vary substantially, from 351 in Alagoas to 426 in the Federal District. This spread of 75 points corresponds to 75% of the individual-level standard deviation in the OECD countries, or about 11 IQ points.

2. Tables 2 and 4 show that in Brazil, the closest correlate of average family income is the percentage of university graduates. Although we would expect a high correlation between PISA scores and higher education, the initially moderately high correlations between these variables tend to fade into insignificance when only neighboring states are compared. High correlations of cognitive measures with higher education have been reported at the subnational level from other countries, for example the association in Turkey between regional IQ and higher education graduation rate (*r*=.63) (Lynn, Sakar & Cheng, 2015). A likely reason for the lack of robusticity in the relationship between PISA scores and higher education in Brazil is that PISA scores reflect mainly the intellectual prowess of the average school child, whereas the proportion of university graduates depends on the size of the intellectual elite. The high correlation between higher education and income may also indicate that universities are more likely to be established in prosperous states rather than poor states.

3. There is a negative correlation of cognitive ability with infant mortality that clearly persisted even in comparisons between neighboring states, suggesting a rather direct relationship between the two variables. Similar associations had been reported before for a number of countries including the regions of the British Isles (*r*=-.78), France (*r*=-.30), Italy (*r*=-.80), the American states (*r*=-.54), Finland (*r*=-.79), India (*r*=-.39), European Russia in the late nineteenth century (*r*=-.28), Finland and the contemporary Russian Federation (*r*=-.43) (Dutton & Lynn, 2014; Grigoriev, Lapteva & Lynn, 2016; Grigoriev, Ushakov

et al., 2016; Lynn, 1979, 1980, 2010; Lynn & Yadav, 2015; Reeve & Basalik, 2011). These results at the aggregate level are consistent with the negative association of infant mortality with IQ of mothers reported by Savage (1946). It is proposed that the explanation for these correlations is that people with higher intelligence are more competent in looking after their babies. They are better able to avoid accidents, and are more likely to be knowledgeable about infant nutrition and hygiene, with the result that their infants are healthier and less likely to succumb to diseases. The association can also be explained by scarcity of resources, since the negative correlation of infant mortality with income and especially its positive correlation with poverty are high (Table 2). The relationship with poverty is still highly significant in the correlation among difference scores (Table 4).

4. The negative correlation of intelligence with fertility is still evident after the spatial autocorrelation control. It confirms the negative correlations of intelligence with fertility rates across the American states (*r*=-.37), the regions of Turkey (*r*=-.89) and India (*r*=-.25), European Russia in the late nineteenth century (*r*=-.28), and the contemporary Russian Federation (*r*=-.39) (Grigoriev, Lapteva & Lynn, 2016; Grigoriev, Ushakov et al., 2016; Lynn, Sakar & Cheng, 2015; Lynn &Yadav, 2015; Shatz, 2009).These results are consistent with studies in many countries showing that dysgenic fertility for intelligence has been near-universal during the twentieth and twenty-first centuries (Lynn, 2011; Woodley & Figueredo, 2013). In Latin America, this has been confirmed through path analysis by León & Avilés (2016), who reported cognitive ability → fertility path coefficients ranging between -.25 and -.40 for a set of more than 1000 districts in Peru.

5. The positive correlation of intelligence with life expectancy was found to be robust even with control for spatial autocorrelation. This result is consistent with a number of studies reporting an association at the individual level between intelligence and longevity, e.g. Deary, Whalley and Starr (2009, p. 50); Gottfredson and Deary (2004).

6. The negative correlations of intelligence with violent crime confirm the negative association in the United States between state IQs and violent crime (*r*=-.58) reported by Bartels et al. (2010) and in Japan between regional IQs and homicide rates (*r*=-.60) reported by Kura (2013). These results are consistent with numerous reports of an association between crime and low intelligence summarized in more than a hundred studies in a number of countries by Ellis and Walsh (2003). However, the state-level correlations we observe between violent crime and PISA scores are rather low and do not survive the spatial

autocorrelation control. Factors other than cognitive ability appear to be the more important determinants of violent crime in Brazil. Violence among human beings can be due to the lack of essential resources (Allen et.al, 2016). However, the weakness of the relationship between income and violent crime (Tables 2, 4, 6) suggests that this may not be of great importance in Brazil—or, perhaps, that mean income is not a good indicator for the condition of poorer people in Brazil, as it is disproportionately determined by the numbers and income of the rich. Median rather than mean income would be a better indicator for the standard of living of average people.

Principal components analysis was performed to obtain conceptual clarity about the socio-economic factors that are most closely related to intellectual ability. This analysis produced three interpretable factors. The first is represented by infant mortality, water availability, poverty, and fertility. It describes the living conditions of the poorer sections of the population. This factor is strongly related to PISA scores as shown in Table 5, perhaps because most of the students taking the PISA tests are the children of relatively poor parents. In addition to their numerical strength, the lower socio-economic groups have higher fertility than wealthier and more educated people, and therefore make a greater contribution to the next generation. For example, in the Brazilian sample of the World Values Survey the correlation between education and fertility is -.229 for females (N = 763) and -.192 for males (N = 495) with completed fertility. As a result, a disproportionate fraction of the school children taking the PISA tests are from lower socio-economic origins.

The second principal component includes income and higher education. The high correlation between these two variables suggests a strong bidirectional relationship, with higher education leading to high income and high income being an important prerequisite for obtaining a university education in Brazil. This relationship is observed on the background of a high level of income inequality, with a Gini index above 50 according to United Nations data (http://www.wider.unu.edu/research/Database/). In consequence, the *average* income most closely reflects the proportion of high earners in a community or state, which in turn is proxied by the proportion of people with a university degree.Thus the income-education factor describes the abundance of wealthy and educated people in the state. This "prosperity factor" is related less closely to PISA scores than is the "poverty factor." The third principal component is a "crime" factor that has relatively weak connections with both PISA scores and with other socio-economic indicators.

Does the relationship of PISA score with the "poverty factor" mean that higher average intelligence and/or education of the general population is required in order to achieve better living conditions for common people and the poor? Or does it mean that the development of higher intelligence, at least when measured as school achievement, depends on decent living conditions? These questions cannot be answered with the present data, which are cross-sectional rather than longitudinal. The first hypothesis predicts that PISA scores today have a direct effect on standards of living 20 or 30 years later, when the children who are tested today are economically active. The second hypothesis predicts that standards of living affect PISA scores (measured at age 15) up to 15 years later. Future studies with longitudinal data will have to address these questions.

Even without a definitive answer to the question of whether intelligence produces wealth or wealth produces intelligence, we can approach the more general question of whether cognitive resources or material resources are more important for various outcomes that people care about. One possibility is that the material resources available to a society, measured as mean income or per capita GDP, are the direct cause for multiple outcomes related to health, standard of living, crime rates and the like. A second possibility is that cognitive resources have direct effects on these outcomes that are not mediated by material wealth alone. The results presented in Table 6 show that in Brazil, cognitive resources do matter for many outcomes at the state level independent of the available material resources, in addition to effects that cognitive development may have on the generation of material resources. One implication is that successful efforts at raising the average cognitive level of the population can produce multiple benefits, even independent of their effects on economic growth. Surprisingly, PISA scores appear to even have a poverty-reducing effect that is independent of *average* income. This may indicate that brighter people are either more able or more inclined than less intelligent ones to redistribute resources in order to restrain poverty when material resources are in limited supply.

It is also evident that the correlations among most of the variables are quite high at the level of Brazilian states, compared to similar correlations for subnational administrative units in other countries. One explanation for the strength of the correlations is the inequality of the income distribution: in Brazil there are states with a level of development similar to poor countries in Asia and Africa, and there are states with a near-European level of development. Another explanation is investment in education. Although primary education is compulsory and free in Brazil, states like Paraná and São Paulo have high levels of school enrolment, while states like Maranhão and Alagoas do not. These inequalities between states invariably lead to high correlations. There are at least two explanations for poor cognitive development in some of the states: (1) the financial distress of the poor leads people to work rather than study, which leads to educational and cognitive deficits; and (2) financial needs restrict access to foods and especially to healthy foods, which leads to poor brain development and cognitive deficits. These possibilities will be investigated in future studies.

References

Allen, M.W., Bettinger, R.L., Codding, B.F., Jones, T.L. & Schwitalla, A.W. (2016). Resource scarcity drives lethal aggression among prehistoric hunter-gatherers in central California. *Proceedings of the National Academy of Sciences* 113: 12120-12125.

Atlas of Human Development in Brazil (2013). Available at: http://www.atlasbrasil.org.br.

Bartels, M., Rietveld, M.J.H., van Baal, G.C.M. & Boomsma, D.I. (2002). Heritability of educational achievement in 12 year olds and the overlap with cognitive abilities. *Twin Research* 5: 544-553.

Bartels, J.M., Ryan, J.J., Urban, L.S. & Glass, L.A. (2010). Correlations between state IQ and FBI crime statistics. *Personality and Individual Differences* 48: 579-583.

Carl, N. (2016). IQ and socioeconomic development across regions of the UK. *Journal of Biosocial Science* 48: 406-417.

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.

Curi, A.Z. & Menezes-Filho, N. (2014).The relationship between school performance and future wages in Brazil. *Economia* 15(3): 261-274.

Da Silva, J.A., Ribeiro-Filho, N.P. & Santos, R.C.D. (2012). Inteligência humana e suas implicações. *Temas em Psicologia* 20(1): 155-188.

Davenport, K.S. & Remmers, H.H. (1950). Factors in state characteristics related to average A-12/V-12 scores. *Journal of Educational Psychology* 41: 110-115.

Deary, I.J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence* 35: 13-21.

Deary, I.J., Whalley, L.J. & Starr, J.M. (2009). *A Lifetime of Intelligence*. Washington DC: American Psychological Association.

Dutton, E. & Lynn, R. (2014). Regional differences in intelligence and their social and economic correlates in Finland. *Mankind Quarterly* 54: 447-456.

Eff, E.A. (2004). Spatial and cultural autocorrelation in international datasets. *Department of Economics and Finance Working Paper Series*, available at http://www.mtsu.edu/~berc/working/spatial%20autocorrelation%20z.pdf

Ellis, L. & Walsh, A. (2003). Crime, delinquency and intelligence. In: H. Nyborg (ed.), *The Scientific Study of General Intelligence*. Amsterdam: Elsevier.

Flores-Mendoza, C., Widaman, K.F., Mansur-Alves, M., da Silva Filho, J.H., Pasian, S.R. & Schlottfeldt, C.G.M.F. (2012). Considerations about IQ and human capital in Brazil. *Temas em Psicologia* 20(1): 133-154.

Folha de São Paulo (2015). Available at: http://www1.folha.uol.com.br/cotidiano/2015/10/1691547-casos-de-morte-violenta-intencional-aumentam-em-18-estados.shtml. Access: 23/07/2016.

Fortin, M.J., Dale, M.R. & VerHoef, J.M. (2002). Spatial analysis in ecology. In: A.H. El Shaarawi & W.W. Piegorsch (eds.), *Encyclopedia of Environmetrics*. Wiley Stats Ref: Statistics Reference Online.

Fuerst, J. & Kirkegaard, E.O. (2016). Admixture in the Americas: Regional and national differences. *Mankind Quarterly* 56: 255-373.

Gottfredson, L.S. & Deary, I.J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science* 13: 1-4.

Grigoriev, A., Lapteva, E. & Lynn, R. (2016). Regional differences in intelligence, infant mortality, stature and fertility in European Russia in the late nineteenth century. *Intelligence* 55: 34-37.

Grigoriev, A., Ushakov, D., Valueva, E., Zirenko, M. & Lynn, R. (2016). Differences in educational attainment, socio-economic variables and geographical location across 79 provinces of the Russian Federation. *Intelligence* 58: 14-17.

Hassall, C. & Sherratt, T.N. (2011). Statistical inference and spatial patterns in correlates of IQ. *Intelligence* 39: 303-310.

Jones, G. (2016). *Hive Mind: How Your Nation's IQ Matters so Much More than Your Own*. Stanford: Stanford University Press.

Kanazawa, S. (2006). IQ and the wealth of states. *Intelligence* 34: 593-600.

Kaufman, S.B., Reynolds, M.R., Liu, X., Kaufman, A.S. & McGrew, K.S. (2012). Are cognitive *g* and academic achievement g one and the same? An exploration of the Woodcock-Johnson and Kaufman tests. *Intelligence* 40: 123-138.

Kirkegaard, E.O. (2015). The S factor in Brazilian states. *The Winnower*.

Kura, K. (2013). Japanese north–south gradient in IQ predicts differences in stature, skin color, income, and homicide rate. *Intelligence* 41: 512-516.

Laidra, K., Pullmann, H. & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences* 42: 441-451.

Léon, F.R. & Avilés, E. (2016). How altitude above sea levels affects intelligence. *Intelligence* 58: 33-41.

Lynn, R. (1979). The social ecology of intelligence in the British Isles. *British Journal of Social and Clinical Psychology* 18: 1-12.

Lynn, R. (1980). The social ecology of intelligence in France. *British Journal of Social and Clinical Psychology* 19: 325-331.

Lynn, R. (2010). In Italy, north-south differences in IQ predict differences in income, education and infant mortality. *Intelligence* 38: 93-100.

Lynn, R. (2011). *Dysgenics: Genetic Deterioration in Modern Populations*, 2nd revised edition. London: Ulster Institute for Social Research.

Lynn, R. (2012). North-south differences in Spain in IQ, educational attainment, per capita income, literacy, life expectancy and employment. Mankind Quarterly *52: 265-291.*

Lynn, R. & Cheng, H. (2013). Differences in intelligence across thirty-one regions of China and their economic and demographic correlates. *Intelligence* 41: 553-559.

Lynn, R., Sakar, C. & Cheng, H. (2015). Regional differences in intelligence, income and other socio-economic variables in Turkey. *Intelligence* 50: 144-150.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. Ulster Institute for Social Research.

Lynn, R. & Yadav, P. (2015). Differences in cognitive ability, per capita income, infant mortality, fertility and latitude across the states of India. *Intelligence* 49: 179-185.

McDaniel, M.A. (2006). Estimating state IQ: Measurement challenges and preliminary correlates. *Intelligence* 34: 607-619.

G. Meisenberg (2014). Cognitive human capital and economic growth in the 21st century. In: T. Abrahams (ed.), *Economic Growth in the 21st Century: New Research*, pp. 49-106. New York: Nova Publishers.

Meisenberg, G. & Lynn, R. (2011). Intelligence: A measure of human capital in nations. *Journal of Social, Political and Economic Studies* 36: 421-454.

Montmollin, M. (1958). Le niveau des recrues du contingent. *Population* 13: 259-268.

O Estado de São Paulo (2016). Available at: http://www.em.com.br/app/noticia/nacional/2016/10/28/interna_nacional,818765/mortes-caem-1-2-no-paislatrocinio-cresce-7-8.shtml. Access: 08/11/2016.

Organization for Economic Cooperation and Development – OECD (2016). Available at: https://www.oecd.org/pisa/pisaproducts/.

Piffer, D. & Lynn, R. (2014). New evidence for differences in fluid intelligence between north and south Italy and against school resources as an explanation for the north-south IQ differential. *Intelligence* 46: 246-249.

Plomin, R., Kovas, Y. & Haworth, C. (2007). Generalist genes: Genetic links between brain, mind, and education. *Mind, Brain and Education* 1(1): 11-19.

Poropat, A.E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin* 135: 322-338.

Reeve, C.L. & Basalik, D. (2011). A state level investigation of the associations among intellectual capital, religiosity and reproductive health. *Intelligence* 39: 64-73.

Rindermann, H. (2007). The *g*-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ tests across nations. *European Journal of Personality* 21: 667-706.

Rindermann, H. & Neubauer, A.C. (2004). Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models. *Intelligence* 32: 573-589.

Rindermann, H., Sailer, M. & Thompson, J. (2009). The impact of smart fractions, cognitive ability of politicians, and average competence of peoples on social development. *Talent Development and Excellence* 1(1): 3-25.

Roivainen, E. (2012). Economic, educational and IQ gains in eastern Germany 1990-2006. *Intelligence* 40: 571-575.

Savage, S.W. (1946). Intelligence and infant mortality in problem families. *British Medical Journal* 19 (Jan): 86-87.

Shatz, S.M. (2009). State IQ and fertility in the United States. *Mankind Quarterly* 49: 38-49.

Snyderman, M. & Rothman, S. (1988). *The IQ Controversy, the Media and Public Policy*. Transaction Publishers.

Sorjonen, K., Hemmingsson, T., Deary, I.J. & Melin, B. (2015). Mediation of the gravitational influence of intelligence on socio-economic outcomes. *Intelligence* 53: 8-15.

Templer, D.I. (2012). Biological correlates of northern–southern Italy differences in IQ. *Intelligence* 40: 511-517.

Woodley, M.A. & Figueredo, A.J. (2013). *Historical Variability in Heritable General Intelligence. Its Evolutionary Origins and Socio-Cultural Consequences*. Buckingham, UK: University of Buckingham Press.

Appendix 1. *PISA total scores and economic and social data for the 27 federative units of Brazil.*

Federative unit	PISA score	Higher educ.	Income	Infant mortality	Water available	Life expect.	Fertility	Poverty	Violence 2014	Violence 2015
Acre	378.1	8.98	522.2	23.0	47.24	71.63	2.95	15.59	26.83	25.3
Alagoas	351.4	6.90	432.6	28.4	75.64	70.32	2.22	16.66	66.48	50.8
Amapá	374.9	10.84	599.0	15.1	66.38	73.80	2.48	9.93	35.56	40.6
Amazonas	379.8	8.23	539.8	17.0	62.16	73.30	2.59	16.43	26.49	37.1
Bahia	370.6	6.40	496.7	21.7	77.60	71.97	2.05	13.79	41.42	41.7
Ceará	385.5	7.60	460.6	19.3	76.28	72.60	1.99	14.69	50.77	46.1
Distrito Federal	425.8	23.95	1715.1	14.0	96.01	77.35	1.75	1.19	25.84	23.4
Espírito Santo	421.4	11.06	815.4	14.2	96.01	75.10	1.80	2.67	42.21	37.4
Goiás	397.8	10.27	810.6	14.0	93.66	74.60	1.87	2.32	42.86	44.3
Maranhão	356.8	5.43	360.3	28.0	51.79	70.40	2.56	22.47	32.17	33.8
Mato Grosso	385.3	10.47	762.5	16.8	90.37	74.25	2.08	4.41	42.64	41.3
Mato Grosso do Sul	405.1	11.99	799.3	18.1	93.76	74.96	2.04	3.55	24.39	22.6
Minas Gerais	417.9	10.57	749.7	15.1	94.91	75.30	1.79	3.49	19.72	20.8
Pará	377.6	6.21	446.8	20.3	57.50	72.36	2.50	15.90	44.81	45.8
Paraiba	386.1	8.02	474.9	21.7	78.91	72.00	1.95	13.39	38.36	37.8
Paraná	416.6	12.75	890.9	13.1	96.69	79.80	1.86	1.96	25.35	25.2
Pernambuco	376.3	8.01	525.6	20.4	78.22	72.32	1.92	12.32	37.02	41.6
Piauí	378.7	7.29	416.9	23.1	67.12	71.62	1.99	18.71	22.92	20.8
Rio de Janeiro	396.1	14.41	1039.3	14.2	94.37	75.10	1.68	1.98	34.74	30.3
Rio Grande do Norte	375.5	8.32	545.4	19.7	85.06	72.52	1.98	10.33	49.99	48.6
Rio Grande do Sul	414.2	11.28	959.2	12.4	96.46	75.38	1.76	1.98	22.16	24.7
Rondônia	387.6	8.04	670.8	18.0	79.62	72.97	2.16	6.39	30.88	31.0
Roraima	376.6	10.16	605.6	16.1	74.04	73.51	2.41	15.66	14.69	18.2
Santa Catarina	418.8	12.53	983.9	11.5	97.00	76.61	1.71	1.01	13.77	14.3
Sao Paulo	409.0	15.10	1084.5	13.9	97.12	75.69	1.66	1.16	12.74	11.7
Sergipe	388.9	8.53	523.5	22.2	82.24	71.84	1.95	11.70	48.93	57.3
Tocantins	374.3	10.30	586.6	19.6	80.41	72.56	2.41	10.21	24.12	25.7

Appendix 2. *Difference scores between each state and its neighbors for PISA and economic and social variables.*

Federative unit	Neighbors	PISA score	Higher educ.	Income	Infant mortality	Water available	Life expect.	Fertility	Poverty	Violence 2014	Violence 2015
Acre	Amazonas, Rondônia	-5.6	0.85	-89.2	5.50	-23.65	-1.51	0.58	4.18	-1.86	-8.75
Alagoas	Bahia, Pernambuco, Sergipe	-27.2	-0.75	-82.7	6.97	-3.71	-1.72	0.25	4.06	24.02	3.90
Amapá	Pará	-2.7	4.63	152.2	-5.20	8.88	1.44	-0.02	-5.97	-9.25	-5.20
Amazonas	Acre, Mato Grosso, Pará, Roraima, Rondônia	-1.2	-0.54	-61.8	-1.84	-7.59	0.36	0.17	4.84	-5.48	4.78
Bahia	Alagoas, Espírito Santo, Goiás, Maranhão, Minas Gerais, Pernambuco, Piauí, Sergipe, Tocantins	-14.2	-2.30	-83.4	1.14	-2.40	-0.70	-0.01	2.62	4.04	4.76
Ceará	Paraíba, Pernambuco, Piauí, Rio Grande do Norte	6.3	-0.31	-30.1	-1.93	-1.05	0.49	0.03	1.00	13.70	8.90
Distrito Federal	Goiás, Minas Gerais	17.9	13.530.	935.0	-0.55	1.73	2.40	-0.08	-1.72	-5.45	-9.15
Espírito Santo	Bahia, Mato Grosso, Rio de Janeiro	26.5	0.600.	53.5	-2.80	7.06	0.98	-0.04	-3.75	10.25	6.47
Goiás	Bahia, Distrito Federal, Mato Grosso, Mato Grosso do Sul, Minas Gerais,Tocantins	1.3	-3.79	0.2	4.82	-3.55	-2.00	-2.00	-2.00	-2.00	-2,002
Maranhão	Bahia, Pará, Piauí, Tocantins	-18.5	-2.11	-126.4	6.83	-18.87	-1.73	0.32	7.82	-1.15	.30
Mato Grosso	Amazonas, Goiás, Mato Grosso do Sul, Pará, Roraima, Tocantins	-1.7	1.31	120.2	-1.03	12.52	0.79	-0.18	-4.72	10.38	6.88
Mato Grosso do Sul	Goiás, Mato Grosso, Paraná, São Paulo	2.9	-0.16	-87.8	3.65	-0.70	-1.13	0.17	1.09	-6.51	-8.03
Minas Gerais	Bahia, Distrito Federal, Espírito Santo Goiás, Mato Grosso do Sul, Rio de Janeiro	14.5	-0.36	-0.4	-0.36	-0.36	-0.36	-0.36	-0.36	-0.36	-0,362
Pará	Amapá, Amazonas, Maranhão, Mato Grosso, Roraima, Tocantins	3.00	-3.02	-128.9	1.53	-13.36	-0.61	0.08	2.72	15.53	13.02
Paraíba	Ceará, Pernambuco Rio Grande do Norte	7.0	0.04	-35.6	1.90	-0.94	-0.48	-0.01	0.94	-7.57	-7.63
Paraná	Mato Grosso do Sul, Santa Catarina, São Paulo	5.6	-0.46	-65.0	-1.40	0.73	4.05	0.06	0.05	8.38	9.00
Pernambuco	Alagoas, Bahia, Ceará, Paraíba, Piauí	1.9	0.77	69.3	-2.44	3.11	0.62	-0.12	-3.13	-6.97	2.16
Piauí	Bahia, Ceará, Maranhão, Pernambuco	6.4	0.43	-43.9	0.75	-3.85	-0.20	-0.14	2.89	-17.43	-20.00
Rio de Janeiro	Espírito Santo, Minas Gerais, São Paulo	-20.0	2.17	156.1	-0.20	-1.64	-0.26	-0.07	-0.46	9.85	7.00
Rio Grande do Norte	Ceará, Paraíba	-10.3	0.51	77.6	-0.80	7.47	0.22	0.01	-3.71	5.43	6.65
Rio Grande do Sul	Santa Catarina	-4.6	-1.25	-24.7	0.90	-0.54	-1.23	0.05	0.97	8.39	10.40
Rondônia	Acre, Amazonas, Mato Grosso	6.5	-1.19	62.7	-0.93	13.03	-0.09	-0.38	-5.75	-1.11	-3.57
Roraima	Amazonas, Pará	-2.2	2.94	112.3	-2.55	14.21	0.68	-0.14	-0.51	-20.96	-23.25
Santa Catarina	Paraná, Rio Grande do Sul	3.4	0.52	58.8	-1.25	0.43	-0.98	-0.10	-0.96	-9.99	-10.65
Sao Paulo	Goiás, Mato Grosso do Sul, Minas Gerais, Paraná, Rio de Janeiro	0.1	2.67	214.7	-1.23	2.19	-0.60	-0.18	-1.59	-13.31	-13.03
Sergipe	Alagoas, Bahia	27.9	1.88	58.9	-2.85	5.62	0.70	-0.19	-3.53	-5.02	11.05
Tocantins	Bahia, Goiás, Maranhão, Mato Grosso, Pará,	-3.4	2.49	11.2	-0.56	6.23	-0.16	0.20	-1.57	-16.66	-15.68

Admixture in Argentina

Emil O.W. Kirkegaard
Ulster Institute for Social Research, London, UK

John Fuerst
Ulster Institute for Social Research, London, UK

*Corresponding author, email: emil@emilkirkegaard.dk

Analyses of the relationships between cognitive ability, socioeconomic outcomes, and European ancestry were carried out at multiple levels in Argentina: individual (max. n = 5,920), district (n = 437), municipal (n = 299), and provincial (n = 24). Socioeconomic outcomes correlated in expected ways such that there was a general socioeconomic factor (S factor). The structure of this factor replicated across four levels of analysis, with a mean congruence coefficient of .96. Cognitive ability and S were moderately to strongly correlated at the four levels of analyses: individual r=.55 (.44 before disattenuation), district r=.52, municipal r=.66, and provincial r=.88. European biogeographic ancestry (BGA) for the provinces was estimated from 25 genomics papers. These estimates were validated against European ancestry estimated from self-identified race/ethnicity (SIRE; r=.67) and interviewer-rated skin brightness (r=.33). On the provincial level, European BGA correlated strongly with scholastic achievement-based cognitive ability and composite S-factor scores (r's .48 and .54, respectively). These relationships were not due to confounding with latitude or mean temperature when analyzed in multivariate analyses. There were no BGA data for the other levels, so we relied on %White, skin brightness, and SIRE-based ancestry estimates instead, all of which were related to cognitive ability and S at all levels of analysis. At the individual level, skin brightness was related to both cognitive ability and S. Regression analyses showed that SIRE had little detectable predictive validity when skin brightness was included in models.

Similarly, the correlations between skin brightness, cognitive ability, and S were also found inside SIRE groups. The results were similar when analyzed within provinces. In general, results were congruent with a familial model of individual and regional outcome differences.
Key Words: Cognitive ability, Socioeconomic outcomes, Inequality, SES, General socioeconomic factor, S-factor, Race, Ethnicity, Ancestry, Biogeographical ancestry, Admixture, Skin color, Argentina.

The causes of differences in social outcomes between countries, and between ethnic groups within countries, represent two topics of active sociological inquiry. Frequently, the two are analyzed independently with, for example, climatic and environmental factors such as UV radiation (León & Burga-León, 2015), cold climate (Van de Vliert, 2013), and parasite load (Eppig, Fincher & Thornhill, 2010) being investigated as causes of the former and with social processes, such as racial discrimination and colorism (Rangel, 2014; Telles, 2014), being investigated as causes of the latter.

Some have proposed and explored a theoretically more parsimonious familial model. According to this model individual-level cognitive ability (CA) or human capital differences are transmitted across generations within families, and are the cause of both inter-country and inter-ethnic differences (Easterly & Levine, 2016; Lynn, 2008, 2015; Putterman & Weil, 2010; Spolaore & Wacziarg, 2013). Because the members of these countries and ethnic groups frequently do or recently did form nations in the traditional sense (Latin *Natio*: stock, race, tribe), this familial model predicts that ancestry will predict, to some degree, both inter-country and inter-ethnic differences within multi-ethnic countries. Support for this familial model was found by Putterman and Weil (2010), who concluded:

The overall finding of our paper is that the origins of a country's population – more specifically, where the ancestors of the current population lived some 500 years ago – matter for economic outcomes today. Having ancestors who lived in places with early agricultural and political development is good for income today, both at the level of country averages and in terms of an individual's position within a country's income distribution. Exactly why the origins of the current population matter is a question on which we can only speculate at this point. People who moved across borders brought with them human capital, cultures, genes, institutions, and languages. People who came from areas that developed early evidently brought with them versions of one or more of these things that were conducive to higher income. Future research will have to sort out which ones were the most significant.

While Putterman and Weil looked at the effect of national-level ancestry on outcomes both between and within countries, Fuerst and Kirkegaard (2016a, 2016b) and Kirkegaard, Wang and Fuerst (2017) examined the effect of continental biogeographic ancestry (BGA) in the Americas. Mirroring Putterman and Weil's results, the authors found that, at both the individual and regional levels, European BGA, relative to African and Amerindian, was a robust predictor of cognitive and/or socioeconomic outcomes (regional: mean correlations of .71 and .64, individual: .18). Furthermore, mediation analyses showed that the relationship between regional European ancestry and better social outcomes was to a large degree mediated by CA.

These findings are congruent with those of a large number of other studies which have found robust and often strong relationships between CA and better social outcomes at both the individual and aggregate levels (Carl, 2016; Herrnstein & Murray, 1994; Kirkegaard, 2014, 2016a; Lynn & Vanhanen, 2012; Strenze, 2007). On the other hand, and in agreement with León & Burga-León (2015) and Van de Vliert (2013), respectively, the authors also found that absolute latitude and/or cold climate were ecological predictors, which were robust to controls for BGA. Given the results, Fuerst and Kirkegaard (2016b) proposed a co-occurring model, and concluded that an accurate estimation of climatic effects necessitates including ancestry as a control variable in analyses to avoid omitted variable bias.

Fuerst and Kirkegaard (2016b) noted that their results needed further replication, such as by analyzing patterns of regional differences within countries not previously studied. While Fuerst & Kirkegaard (2016a) included Argentina as a country in their pan-American analysis, they did not conduct a regional analysis. Argentina, though, is suitable for this type of analysis, given the reasonable number of major administrative units (24, including the capital district), the size of the population (about 43 million as of 2015), and the diversity in ancestral backgrounds. The primary purpose of this study was to examine how CA, socioeconomic outcomes, and indexes of ancestry relate to each other in Argentina, on both the individual and administrative unit level. The predictions of the specific model being tested are that European ancestry and indicators of this, including skin brightness and self-identified race/ethnicity (SIRE), will be positively associated with cognitive and socioeconomic outcomes. Noting a dearth of data on genetic ancestry, Fuerst and Kirkegaard (2016b) suggested that SIRE and skin brightness data from the Latin American Public Opinion Project (LAPOP; http://www.vanderbilt.edu/lapop/) and related surveys could be used in lieu of genetic data. Thus, a secondary aim of this study was to determine whether LAPOP CA, S, SIRE, and skin brightness data can be used to test the predictions of the familial model. In this regard, it is expected that the ancestry, CA, and S indices constructed from the LAPOP surveys will correlate with ones constructed from other sources.

1. Individual level analysis

We begin by presenting the individual-level analyses. This is because some of the aggregate-level analyses rely on aggregated data from the individual-level analyses. All analyses were carried out in R. The scripts can be found in the supplementary materials. Data sources are discussed below.

1.1. Cognitive variable

While the LAPOP surveys contain no traditional IQ test items (e.g., Raven's figures or synonym questions), individuals are asked several geopolitical knowledge questions, which can be used to create knowledge factor scores. The questions are as follows:

While the LAPOP surveys contain no traditional IQ test items (e.g., Raven's figures or synonym questions), individuals are asked several geopolitical knowledge questions, which can be used to create knowledge factor scores. The questions are as follows:

gi1: What is the name of the president of the United States?
gi2: What is the name of the president of the Senate of Argentina?
gi3: How many provinces does Argentina have?
gi4: How long does the presidential term last in Argentina?
gix4: On which continent is Nigeria located?
gi5: What is the name of the president of Brazil?

The answers to some of the questions vary over time e.g., the president of the US was not the same person in 2008 (George Bush) as in 2014 (Barack Obama). It is possible that this changed the difficulty and discriminative ability of the items. Not all items were given in all surveys. Table 1 shows the item pass rates by year.

Table 1. *Item Pass Rates by Year.*

Year	gi1	gi2	gi3	gi4	gix4	gi5	mean
2008	0.83	0.39	0.50	0.86		0.80	0.67
2010	0.80		0.58	0.75			0.71
2012	0.76			0.86			0.81
2014	0.77			0.78	0.47		0.67

Except for gi2, all items had pass rates >.50, which indicates that there was a lack of difficult items. Scoring such data so that it is comparable across survey years is riddled with difficulty. For instance, a simple method such as taking the mean score for each participant is problematic, as someone who participated in the 2012 survey had to answer two items with a mean pass rate of .81, while a person who participated in the 2008 survey had to answer 5 items with a mean pass rate of .67. Thus, a mean score of .50 would not represent the same level of ability for the two different years. Also, due to the distribution of items across years, some item pairs never appear together. The correlation matrix, calculated using *hetcor* from polycor package (Fox, 2016), is shown in Table 2.

Table 2. *Item correlation matrix using latent correlations.*

	gi1	gi2	gi3	gi4	gi5	gix4
gi1	1.00	0.64	0.53	0.55	0.65	0.60
gi2	0.64	1.00	0.62	0.52	0.59	
gi3	0.53	0.62	1.00	0.48	0.47	
gi4	0.55	0.52	0.48	1.00	0.60	0.42
gi5	0.65	0.59	0.47	0.60	1.000	
gix4	0.60			0.42		1.00

The holes in the correlation matrix mean that one cannot factor analyze it. One solution to this problem is to combine items to remove these holes. Notice in Table 2 that items *gi5* and *gix4* never appear together, but *gi5* appears together with every other item. Because of this, we can combine these two items without any data loss and thereby remove all the holes in the matrix. (This approach is not quite optimal because it combines data from items with dissimilar pass rates of (.47 vs. .80) and somewhat different inter-item correlations (.65 vs. .60; .60 vs. .42).) This results in a new correlation matrix, which is shown in Table 3.

Table 3. *Latent correlations after item combination.*

	gi1	gi2	gi3	gi3	gi45
gi1	1.00	0.64	0.52	0.58	0.65
gi2	0.64	1.00	0.62	0.52	0.59
gi3	0.52	0.62	1.00	0.62	0.47
gi3	0.58	0.52	0.62	1.00	0.60
gi45	0.65	0.59	0.47	0.60	1.00

As expected, the correlations between the new item, *gi45*, are similar to those for the original items (*gix4* and *gi5*). The new dataset was analyzed using item response theory factor analysis. The 2-parameter normal (2PN) model was used. In this model, item responses are modeled as being a function of item difficulties, item discriminative abilities, and each person's latent ability score. There are both more and less complicated models, but this one is the default used by the psych package employed (Revelle, 2015). The advantage of this method is that it takes into account the relative discriminative power (akin to factor loadings) and difficulty (as indicated by pass rates) of items. For instance, a person answering 2 hard items correctly is given a higher score than a person answering 2 easy items correctly (DeMars, 2010; Revelle, 2016). Figure 1 shows the discriminative ability of items when the entire dataset is analyzed together. All items were useful measures of latent CA, with *gi2* being the best. The lack of items on the right side means that the test has limited ability to discriminate between persons with above average CA.

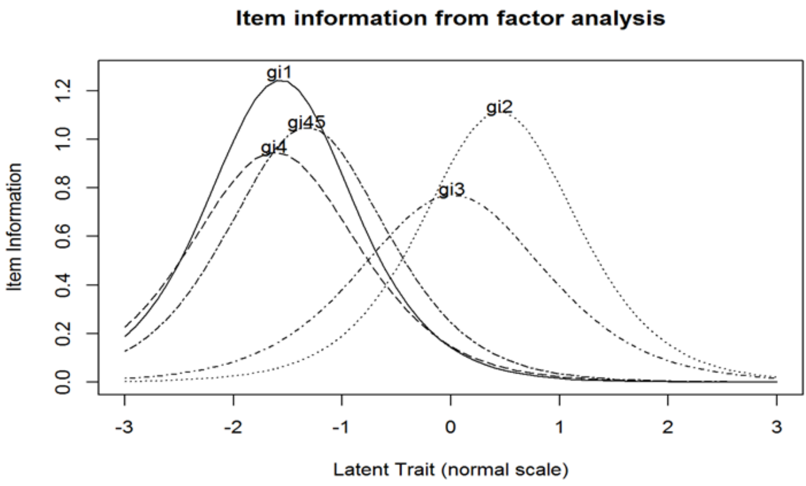

Figure 1. *Discriminative ability of cognitive items.*

Combining individual scores across years was difficult because of the uneven distribution of items. After trying a number of methods, we found that analyzing the data within years, treating missing data as incorrect answers, and standardizing the scores within survey years produced the best results. We experimented with methods by using the S scores (discussed in the next section) as the criterion variable and by seeing which method resulted in the highest criterion correlations. The benefit of standardizing the data within survey years is that each survey has the same mean and standard deviation despite the uneven item distribution. Still, the combined score distribution had a long left tail (skew: -.79). This was mostly due to the lack of harder items and to the extremely easy item set used in the 2012 survey.

1.2. Socioeconomic variables

The LAPOP questionnaires asked numerous questions about persons' socioeconomic outcomes. From those, a set of 16 variables was identified:

- Whether they feel safe [Likert]
- Life satisfaction [Likert]
- Educational attainment [number of years]
- Family income [continuous]
- Victim of crime in the last 12 months [dichotomous]
- Ownership or availability of:
    - TV [dichotomous]
    - Refrigerator [dichotomous]
    - Landline phone [dichotomous]
    - Cell phone [dichotomous]
    - Cars [number]
    - Washing machine [dichotomous]
    - Microwave [dichotomous]
    - Motorcycle [dichotomous]
    - Drinking water in the home [dichotomous]
    - Indoors bathroom [dichotomous]
    - Computer [dichotomous]

The variables covered a fairly broad spectrum of socioeconomic outcomes, though some topics were not covered, such as crimes committed and health. About 3% of cells contained missing data. These were imputed using the **VIM** package without noise (single imputation; Templ et al., 2015). Since some of the items were dichotomous (owns a TV, phone etc.), it is preferable to analyze them using a correlation matrix composed of latent (for pairs with one or two dichotomous indicators) and Pearson correlations (for pairs with two continuous indicators). However, using latent correlations means that one cannot regress out the effect of age/sex/survey year on the indicators, though one can do this for the final composite score (S score). The downside to this is that if age/sex/survey year have differential impacts on the indicators, this will not be adjusted for.

Because method variance may be large, we tried both methods, that is, using regular factor analysis on the age/sex/survey year controlled indicators, and using latent correlations + IRT on the discrete data, and then controlling the S scores for age/sex/survey year.

To control for age/sex/survey year, we first fit a local regression (LOESS) model for age to take non-linear effects into account. The residuals were then regressed on sex and survey year in an OLS model. The residuals from the second model were saved for further analysis. The reason to control for survey year is that ownership of some items became more common over time (e.g., computer) for persons of the same socioeconomic level. Figure 2 shows the loadings for the two analyses.

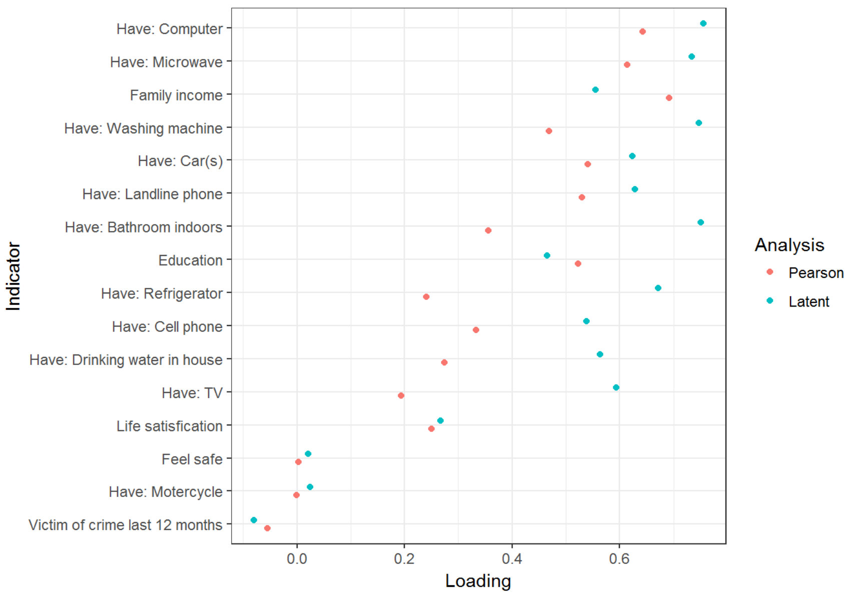

Figure 2. *Socio-economic (S) factor loadings.*

The loadings were stronger when using latent correlations. This is expected because the use of Pearson correlations results in a downwards bias due to dichotomization. The relative order of variables was quite similar, with a factor congruence coefficient of .94 (Lorenzo-Seva & Ten Berge, 2006). Most loadings were in the expected direction. However, feeling safe in one's residential area and being a victim of crime had near-zero loadings. Based on the results above, we decided to use the Pearson-based, age/sex/survey year corrected scores for further analysis.

1.2.1. Robustness checks

Usually, factor analytic methods give near identical results, but in some rare cases they diverge substantially. To check whether this was the case, we factor analyzed the data using all possible combinations of method parameters (30 combinations) (Kirkegaard, 2016b). The median factor score correlation across these was .999 indicating negligible method variance. A general factor extracted from a dataset will be 'flavored' if the indicators in the sample are not representative of all possible indicators (Carroll, 1993; Jensen, 1998). One can measure the extent to which the indicator sampling impacts the scores by splitting the indicators into two subsets at random, extracting the factors in each subset and correlating the scores (Kirkegaard, 2016b). This was done 100 times. The median intercorrelation was .66, indicating that it mattered to some degree which indicators were used for the analysis, but not to the point where using the present indicators would be indefensible. It is possible that the factor structure substantially varies by year, perhaps due to changing social conditions. Thus, as a robustness check, we analyze the S factor structure within years as well. The factor loadings are shown in Figure 3. As can be seen, there was very little variation between years. The mean factor similarity coefficient was .98.

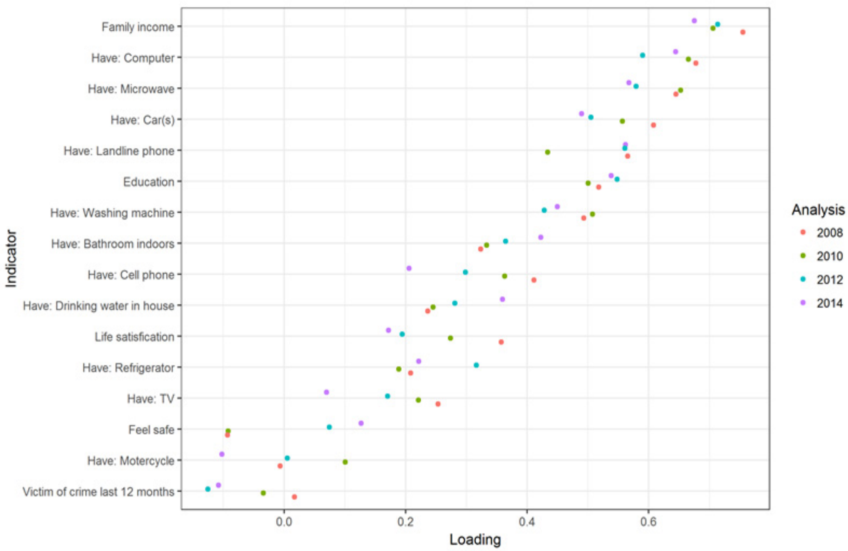

Figure 3. *Socio-economic indicator loadings across years.*

1.3. Demographic variables

The LAPOP questionnaire contained two relevant demographic variables:

SIRE: Blanca (White), Mestiza (Mestizo/Mixed White-Indigeneous), Indígena (Indigeneous), Negra (Negro/Black), Mulata (Mulatto/Mixed White-Black), Otra (Other).

Interviewer-rated skin brightness: 0-9 scale. (We reversed the scoring so that 9 = lightest.)

The relationships between SIRE and skin brightness is shown in Table 4. Despite the small sample sizes in some cases, the order was as one would expect. Whites were the most light-skinned (7.93). Mestizos and Mulattoes fell intermediate (6.73 and 6.29, respectively) along with "Others" (who presumably are mostly admixed East Asians; 6.55). And finally, Indigenous (5.73) were darker, followed by Blacks (4.73).

Table 4. *Summary statistics for skin brightness by self-identified race/ethnicity (SIRE) group.*

SIRE	n	Mean	SD
Blanca	2845	7.93	1.26
Indigena	51	5.73	1.64
Mestiza	1246	6.73	1.48
Mulata	7	6.29	2.29
Negra	37	4.73	1.61
Otra	60	6.55	1.29
(not stated)	173	7.25	1.66

1.4. Relationship between CA and S

We expected there to be a strong relationship between cognitive ability (CA) and socioeconomic status (S). This relationship should be stronger for the years with higher quality CA data. The CA scores for 2008 were based on 5 items, while those from 2012 were based on only 2 easy items. Figure 4 shows a scatterplot of the relationships between CA and S by year. As can be seen, the regression line was very similar across years. This does not necessarily mean that the correlations were also similar, however. Table 5 shows the correlations between the two variables by year. The size of the relationship is worth noting. Despite being based on a 5-item test about political knowledge, the correlation between CA and S still reached .52 in 2008. The correlations were lower for the other years as expected due to the fewer items used, but even using 2 items resulted in a .33

correlation. We estimated test reliability from the IRT analyses and used this to calculate the estimated correlation between CA and S (CA * S) if there was no measurement error in CA. As expected, the reliabilities of the tests were fairly low, especially for the 2-item version (.43). When corrected for the estimated measurement error, the true score correlations were fairly strong, r's .50 to .64. The mean true score correlation (CA * S true) was .55, which was the same as the estimate from the pooled data when that was corrected for measurement error (based on the mean reliability).

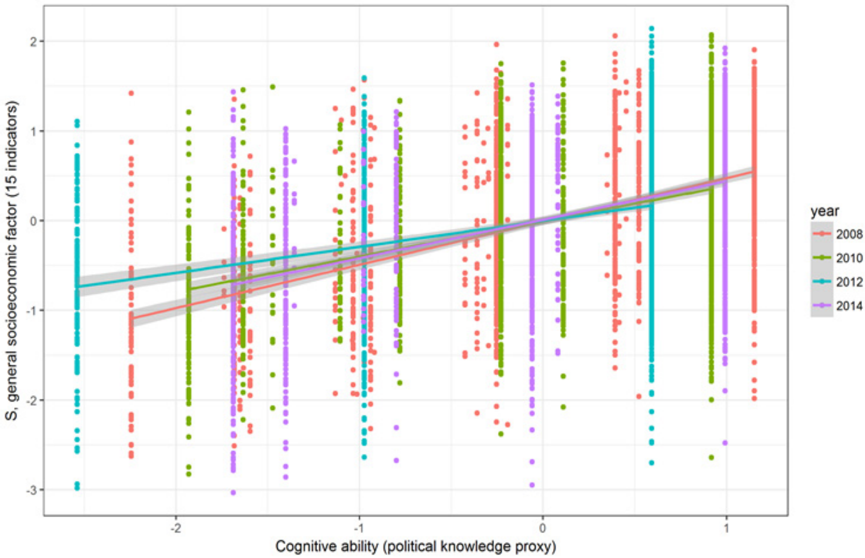

Figure 4. *Scatterplot of cognitive ability and socio-economic (S) factor. Total n = 3,902. Error bands = 95% analytic confidence intervals.*

Table 5. *Correlations between cognitive ability (CA) and socio-economic status (S) by Year.*

Year	CA * S	Reliability	CA * S true
2008	0.52	0.67	0.64
2010	0.44	0.69	0.53
2012	0.33	0.43	0.50
2014	0.48	0.82	0.53
all	0.44	0.65	0.55

1.5. Relationship with demographic values

As mentioned in the introduction, it is well established that SIRE and measures of racial ancestry are associated with both CA and S. A proposed conceptual path model of these relationships is shown in Figure 5.

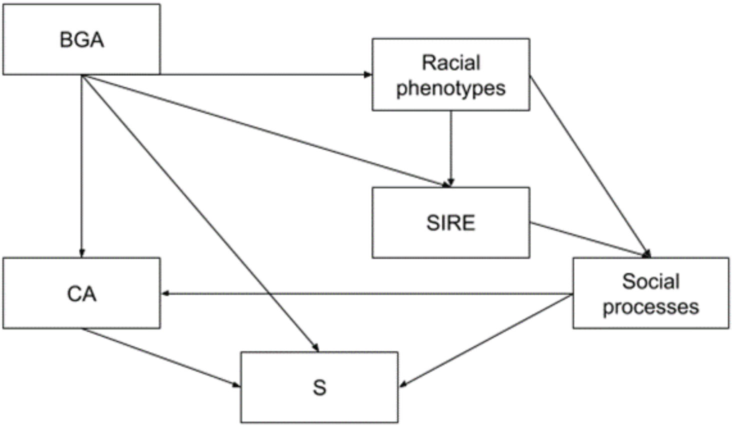

Figure 5. *Conceptual path model.*

In this model, factors associated with BGA are the ultimate (uncaused) cause of social inequality (S) through three proposed mechanisms:

1. Through cognitive ability (BGA→CA→S path).
2. Through other human capital traits that aren't cognitive ability such as personality (BGA→S path).
3. Through racial phenotype-based social processes of favoritism and discrimination (paths involving *Social processes* node).

While (1) and (2) are human capital models based on meritocracy, (3) is not. It would be possible to disentangle the causal network if one had cognitive, socioeconomic, racial phenotype, and genomic ancestry variables. However, the present dataset lacks the fourth. As such, social processes models cannot be disentangled from human capital ones. Nonetheless, SIRE mediated social processing models, specifically, can be tested by including SIRE and racial phenotype jointly as predictors of CA and S. If there are social processes related to SIRE independent of racial phenotypes, SIRE should have incremental predictive validity (assuming no measurement error in the other predictors;

Westfall & Yarkoni, 2016). If, on the other hand, SIRE is a non-causal proxy for factors related to racial phenotype and genetic ancestry, SIRE should have none.

1.6. Relationship between skin brightness, SIRE, CA, and S

There are two main methods one can use to analyze the relative predictive validity of skin brightness and SIRE. First: subsetting the data by SIRE and examining the relationships within each subset. Second: including both in a regression model. The correlations between the variables across all years are shown in Table 6. As expected, skin brightness was weakly to moderately associated with both CA and S, as had been found in other studies (Lynn, 2002). This finding is consistent with both classes of models. If such relationships are due to social processes related to SIRE, they should disappear when data are analyzed within SIRE groups. Table 7 shows the within SIRE correlations.

Table 6. *Correlations between cognitive ability, S, and skin brightness. Correlations corrected for measurement error in cognitive ability are shown above the diagonal, with reliabilities shown in the diagonal.*

	CA	S	Skin brightness
CA	0.65	0.55	0.20
S	0.44	1.00	0.26
Skin brightness	0.16	0.26	1.00

Table 7. *Correlations [with 95% confidence intervals] between cognitive ability (CA), socio-economic status (S) and skin brightness (SB) within SIREs. Correlations were corrected for measurement error in cognitive ability.*

SIRE	SB * CA	SB * S	CA * S	n
Blanca	0.21 [0.17 - 0.24]	0.20 [0.17 - 0.24]	0.55 [0.53 - 0.57]	3869
Indigena	-0.09 [-0.36 - 0.19]	0.34 [0.07 - 0.56]	0.58 [0.41 - 0.71]	77
Mestiza	0.12 [0.06 - 0.17]	0.22 [0.17 - 0.27]	0.52 [0.48 - 0.55]	1596
Mulata	-0.29 [-0.86 - 0.59]	0.65 [-0.21 - 0.94]	-0.21 [-0.65 - 0.34]	15
Negra	0.03 [-0.30 - 0.35]	0.09 [-0.24 - 0.41]	0.38 [0.12 - 0.59]	54
Otra	0.20 [-0.05 - 0.43]	0.49 [0.26 - 0.66]	0.47 [0.27 - 0.63]	73
(not stated)	0.22 [0.08 - 0.36]	0.35 [0.21 - 0.47]	0.56 [0.47 - 0.65]	236

Only three of the SIRE groups had sufficient data to allow for a reasonable estimation of effect size. Within all three, the relationships between skin brightness and the two outcomes were positive, as was the relationship between CA and S. We furthermore calculated the weighted mean correlations across

SIRE groups. These were .18, .22 and .54 for SB * CA, SB * S, and CA * S, respectively. Thus, these relationships were not due to confounding with SIRE or any associated social processes. To establish the baseline validity of SIRE, a model with only that predictor was fit. Secondly, to examine whether SIRE and SIRE-related effects had incremental validity above skin brightness, a model was fit with both predictors. Both models were fit with both CA and S as outcome variables. Finally, a model was fit with S ~ SIRE + skin brightness + CA. In this last scenario, skin brightness reflects social processing effects related to racial phenotypes and non-cognitive ability human capital traits. Unfortunately, due to the large measurement error in the CA variable, the results are not strongly informative because measurement error in one predictor results in its validity being spread to correlated predictors (Westfall & Yarkoni, 2016). Results are shown in Tables 8-12. The R^2 values reported, denoted as R^2-cv, for the regression models were obtained by 10-fold cross-validation to protect against overfitting (James et al., 2013).

Table 8. *Regression Model for cognitive ability with dummy-coded SIRE groups as a predictor. Blanca is the omitted control. N = 5,920; R^2 = .017; R^2-cv = .014; CI = 95% analytic confidence intervals.*

Predictor	Beta	SE	CI lower	CI upper
SIRE: Indigena	-0.56	0.11	-0.78	-0.33
SIRE: Mestiza	-0.15	0.03	-0.21	-0.09
SIRE: Mulata	-0.66	0.26	-1.17	-0.16
SIRE: Negra	-0.68	0.14	-0.95	-0.42
SIRE: not stated	-0.38	0.07	-0.51	-0.25
SIRE: Otra	-0.34	0.12	-0.57	-0.11

Table 9. *Regression model for cognitive ability with dummy-coded SIRE groups and skin brightness as predictors. Blanca is the omitted control. N = 4,419; R^2 = .032; R^2-cv = .027; CI = 95% analytic confidence intervals.*

Predictor	Beta	SE	CI lower	CI upper
SIRE: Indigena	-0.27	0.14	-0.55	0.00
SIRE: Mestiza	0.03	0.04	-0.05	0.10
SIRE: Mulata	-0.57	0.37	-1.30	0.16
SIRE: Negra	-0.19	0.17	-0.52	0.14
SIRE: not stated	-0.29	0.08	-0.44	-0.14
SIRE: Otra	-0.24	0.13	-0.50	0.01
Skin brightness	0.15	0.02	0.12	0.18

By comparing Tables 8 and 9 it can be seen that SIRE alone has some validity (6/6 CIs don't overlap 0), which corresponds to a combined standardized beta of .13. (These equivalent effect sizes were calculated by converting the regression into an analysis of variance, and then calculating eta squared for the predictors. Eta squared corresponds to r^2, and thus one can take the square root to get a measure akin to the standardized beta.) However, when SIRE was entered together with skin brightness, it had little detectable validity (5/6 CIs overlap 0) with a combined effect size equivalent to a standardized beta of .08.

Table 10. *Regression model for socio-economic status (S) with dummy-coded SIRE groups as predictor. Blanca is the omitted control. N = 5,920; R^2 = .043; R^2-cv = .040; CI = 95% analytic confidence intervals.*

Predictor	Beta	SE	CI lower	CI upper
SIRE: Indigena	-0.71	0.11	-0.93	-0.49
SIRE: Mestiza	-0.39	0.03	-0.45	-0.34
SIRE: Mulata	-0.43	0.25	-0.93	0.06
SIRE: Negra	-0.74	0.13	-1.00	-0.47
SIRE: not stated	-0.45	0.07	-0.57	-0.32
SIRE: Otra	-0.45	0.12	-0.67	-0.22

Table 11. *Regression model for socio-economic status (S) with dummy-coded SIRE groups and skin brightness as predictors. Blanca is the omitted control. N = 4,419; R^2 = .074; R^2-cv = .071; CI = 95% analytic confidence intervals.*

Predictor	Beta	SE	CI lower	CI upper
SIRE: Indigena	-0.14	0.14	-0.41	0.13
SIRE: Mestiza	-0.13	0.04	-0.20	-0.06
SIRE: Mulata	0.15	0.36	-0.56	0.87
SIRE: Negra	-0.19	0.16	-0.51	0.13
SIRE: not stated	-0.19	0.08	-0.34	-0.04
SIRE: Otra	-0.23	0.13	-0.48	0.02
Skin brightness	0.24	0.02	0.21	0.27

Likewise, by comparing Tables 10 and 11 it can be seen that SIRE has some validity for S (5/6 CIs don't overlap 0, equivalent to a std. beta of .21), but much less when entered together with skin brightness (4/6 CIs overlap 0, equivalent std. beta of .07). The two SIRE effects that kept some detectable validity had their betas strongly reduced (by 67% and 58%, respectively for Mestiza and 'not stated'). Finally, in Table 12, it can be seen the results hold when CA is also entered as a predictor. A pure CA human capital model predicts the effect of skin brightness should be zero when CA is controlled. However, as mentioned before, the CA measure was poor, which results in imperfect control and thus some residual validity of non-causal correlates is expected. In this case, the validity of skin brightness was reduced by 25% (.24 to .18). As above, Mestiza continued to not overlap with 0. Aggregating over the different SIREs, the effect size attributable to SIRE, skin brightness and CA was equivalent to correlations of .06, .16 and .38, respectively. Thus, whatever social processes related to SIRE exist, their direct effect on social inequality is very small.

Table 12. *Regression model for socio-economic status (S) with dummy-coded SIRE groups, skin brightness and cognitive ability as predictors. Blanca is the omitted control. N = 4,419; R^2 = .220; R^2-cv = 0.210; CI = 95% analytic confidence intervals.*

Predictor	Beta	SE	CI lower	CI upper
SIRE: Indigena	-0.03	0.13	-0.28	0.22
SIRE: Mestiza	-0.14	0.03	-0.20	-0.08
SIRE: Mulata	0.37	0.34	-0.28	1.03
SIRE: Negra	-0.11	0.15	-0.41	0.18
SIRE: not stated	-0.08	0.07	-0.21	0.06
SIRE: Otra	-0.13	0.12	-0.36	0.09
Skin brightness	0.18	0.01	0.15	0.21
CA	0.39	0.01	0.36	0.41

1.7. Robustness check: Within-province relations between skin brightness and outcomes

As a robustness check, we calculated the within province correlations for skin brightness, CA, and S. The correlations were somewhat smaller, with weighted mean correlations of .14, .19 and .51 for CA * SB, S * SB, and CA * S, respectively (corrected for measurement error). These reductions do not necessarily mean that provincial effects were responsible because there was likely reduced variation within provinces for the variables. This analysis was not a primary interest and so the matter was not explored further, though more detailed results can be found in the supplementary files.

2. Province-level analyses

Argentina has 24 provinces. LAPOP reports the province for each person, so it is possible to aggregate the individual-level data by province to form additional province-level variables. In the following, as a weighting scheme, we used the square root of the province's population size unless otherwise noted. For a statistical justification for this weighting scheme, see Fuerst and Kirkegaard (2016a). Results based on other weighting schemes can be found in the supplementary files.

2.1. S factor analysis

As in previous studies (Fuerst & Kirkegaard, 2016a; Kirkegaard, 2016a), we collected a broad set of socioeconomic variables concerning provinces from a statistical agency and subjected them to factor analysis. This was done both to see whether it was empirically possible to speak about provinces with generally better or worse social outcomes (i.e., whether the S factor exists in the data), and because it is convenient to analyze a single aggregate socioeconomic variable instead of many (as with the Human Development Index).

2.1.1. Knoema data

Province-level socioeconomic data were copied from several sources. Most data came from Knoema's Argentina Regional Dataset, October 2013. See datafiles for details. To validate the S-factor scores, Human Development Index (HDI) scores were used. HDI scores for 1996 and 2011 were copied from Argentina, P.N.U.D. (2013). Additional S data were used from the LAPOP surveys.

Including many very similar or even identical variables in a factor analysis tends to 'color' the general factor by these variables. To guard against this, an algorithm was used that excluded variables from the dataset until no pair of variables correlated at -.9 > x >.9. This resulted in the exclusion of 2 variables.

Factor analytic method variance was analyzed by extracting a single factor using 30 combinations of scoring and estimation methods. There was near-zero method variance because the scores from these 30 methods had a mean correlation of .99. Factor analytic indicator sampling error was examined using the split-half method, see Kirkegaard (2016b). The analysis indicated only a small amount of indicator sampling error (median absolute correlation between scores = .85). The primary factor analyses were carried out using Bartlett's scoring method, and combinations of weighted (by population size)/unweighted and interval/rank-level data. The factor loadings are shown in Figure 6. As can be seen, the pattern of loadings was robust across analyses (median factor similarity congruence coefficient = .98). Scores were extracted from the weighted interval-level analysis, as this was the theoretically superior method.

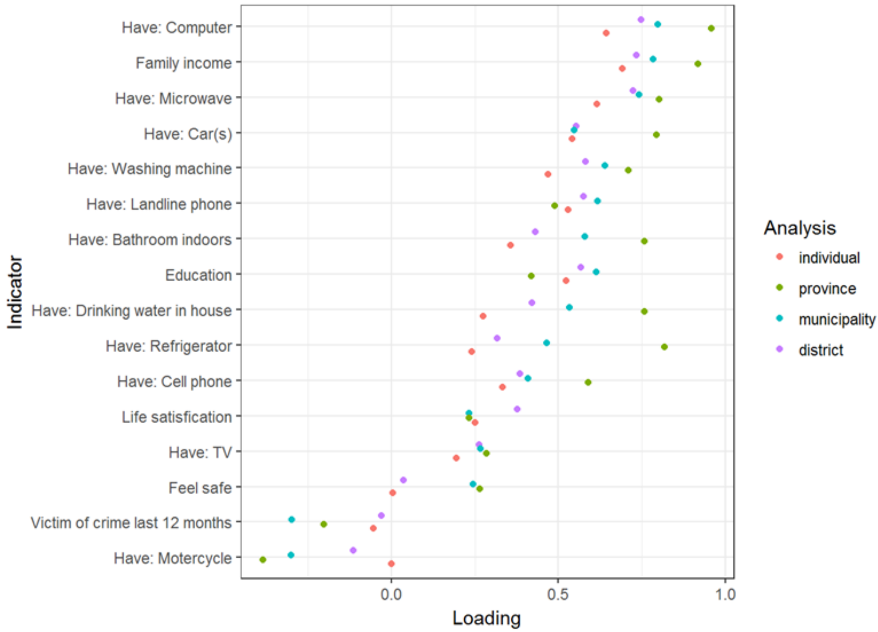

Figure 6. *Factor loadings for the socio-economic (S) factor in the province-level analysis. N = 24 provinces. Based on unweighted values.*

2.1.2. LAPOP data.

There are two ways to utilize the LAPOP data to calculate province-level S factor scores. First, one can factor analyze the data at the individual level, and then average the scores within provinces. Second, one can average the indicator scores within provinces, and then factor analyze the data at the province-level. There is no guarantee that these methods will produce similar results (ecological fallacy/Simpson's paradox; Kievit et al., 2013), however, previous studies have found that they did (Kirkegaard, 2015b, 2016a). The second method has the bonus that one can combine the indicators from it with those from Knoema to produce a larger province-level dataset of indicators. Therefore, indicators were averaged within province and the dataset factor analyzed. We also averaged the indicators for other administrative divisions for comparison (districts and municipalities), which will be discussed later. The factor structure was very similar across levels with a median factor congruence coefficient of .97. In general, loadings were stronger at the aggregate level.

2.1.3. Combined S factor

Given the lack of irregularities, a composite dataset was created from the Knoema indicators and the LAPOP indicators. The loadings from this analysis are shown in Figure 7. Note that because LAPOP had no data for the Santa Cruz province, these data were imputed based on the data available from Knoema. As before, there were no surprising loadings. Crime rate's loading was positive as has been found previously when aggregate data with large units were analyzed (Kirkegaard, 2015a; Lynn, 1979). Curiously, though, at the same time crime victimization was negative. As additional SES measures, we include Human Development Index (HDI) scores for 1996 and 2011. We had previously found that HDI and S factor scores correlated at typically >.90.

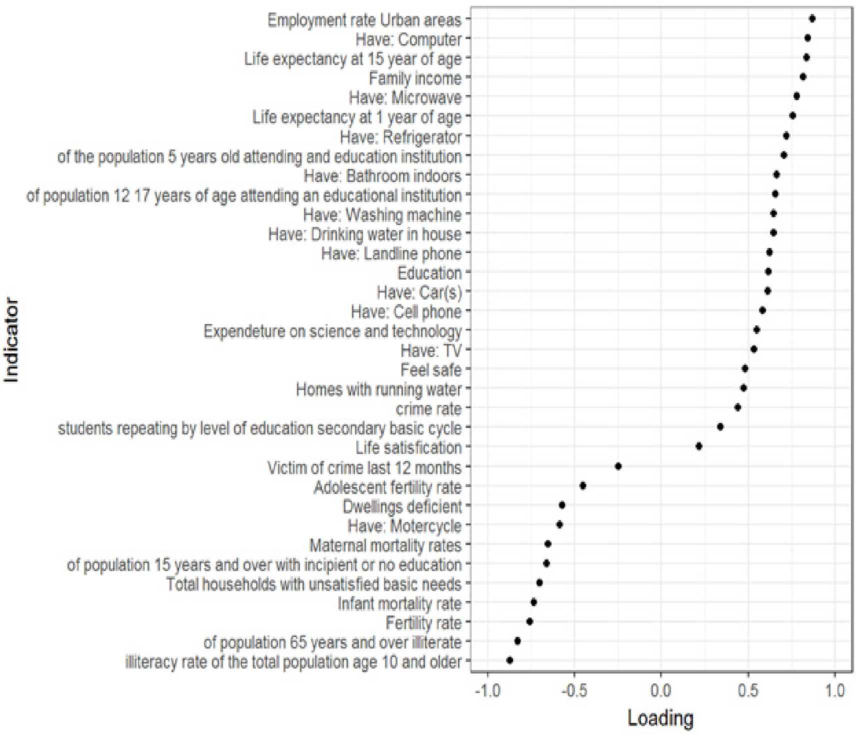

Figure 7. *Socio-economic (S) factor loadings for the combined dataset analyzed at the province level, N = 24 provinces. Based on unweighted values.*

2.2. Cognitive data.

The analysis of the cognitive data is similar to that of the socioeconomic status above in that we have a province-level dataset as well as the ability to create aggregate-level datasets from the LAPOP data. We used two cognitive data sources, Operativo Nacional de Evaluación (ONE) achievement scores and LAPOP political knowledge scores.

As for the ONE scores, we used those from the math, reading, social science and natural science tests administered in 2005, 2006, 2007 and 2010. The 2005 ONE data were based on 6th and 9th graders (primary and secondary school students), while the 2007 and 2010 ONE data were based on 9th graders (secondary school students). Results can be accessed at the Dirección Nacional de Información y Estadística de la Educativa website: http://portales.educacion.gov.ar/diniece/2014/05/22/evaluacion-de-la-calidad-educativa-documentos/. The data provided are unfortunately not mean or median scores, but proportions for three groups: low scoring, medium scoring, and high scoring. These are non-linear transformations of the underlying continuous data (La Griffe du Lion, 2001). We estimated a single best score for each unit by calculating the Z-scores within each combination of grade, year, and subject (32 subdatasets in total), for low and low+medium scoring proportions. The reason to use the combined low and medium group instead of the medium group is that the proportion of students in the medium group is not a good measure of mean level of achievement. This is because a small medium group can mean either that there were many students in the high scoring group or in the low, or in some combination (bimodal). The reason not to use the high scoring group proportion is that it is perfectly negatively correlated with the low+medium proportion and thus redundant. Since it is worse to have more pupils in the lower categories, we reversed the scores. Finally, we averaged these scores to produce a single best estimate for each province across years, grades, and subjects. The median correlation across subdatasets was .86.

As with S scores, there are two ways of estimating mean scores for provinces based on the LAPOP data. First, one can score the persons and then calculate the mean score by province. Second, one can calculate the mean item score by province, factor analyze the data at the province-level and then score the provinces. In this case, the second option should be better because it gets around the problem of having to score persons based on only 2 tests. (As before, we tried both methods to estimate the method variance.)

2.3. Demographic variables

With the aid of a native Argentinean, 25 studies reporting provincial and regional European, African and Amerindian biogeographic ancestry (BGA) were located. Provincial ancestry percentages were calculated as detailed in Supplementary File 1. When provinces had multiple data points, median values were used. No data for either Tierra del Fuego or Entre Rios were available. For these provinces, values were estimated based on the average values of the bordering regions. Figure 8 shows the spatial map of European genetic ancestry. It needs to be noted that these estimates should be viewed with caution, since they are based on studies which often had small, unrepresentative samples, and since estimates were often based on few markers; the latter point is relevant since the number of markers used has been found to substantially influence admixture estimates (Russo et al. 2016).

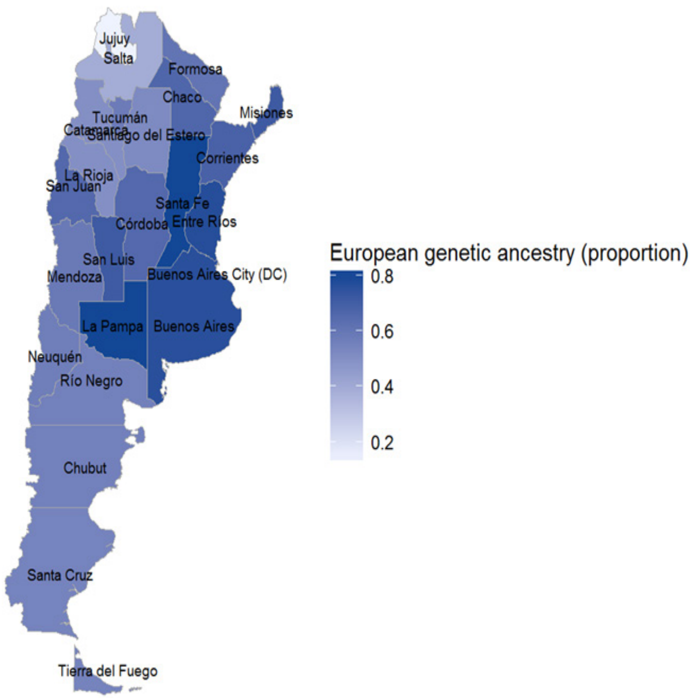

Figure 8. *Map of European genetic ancestry.*

It has been previously found that SIRE-based ancestry estimates correlate strongly with genetic-based ones in the case of Brazilian states and all American nations (Fuerst & Kirkegaard, 2016a). As such, SIRE-based provincial ancestry estimates were computed using self-reported race/ethnicity (White, Black, Indigenous, Mestizo, and Mulatto) following the same procedure as before (Fuerst & Kirkegaard, 2016a). The formula was:

SIRE-based European = %White + ½ %Mulatto + ½ %Mestizo

The SIRE data came from the LAPOP surveys and were based on over 5,800 individuals, aged 16 to 89. No data were available for Santa Cruz, so $n_{European\_SIRE} = 23$.

Fuerst & Kirkegaard (2016a) likewise found that skin reflectance strongly correlated with national European ancestry estimates. Those findings were replicated using national LAPOP interviewee color scores (results not shown). Thus, provincial color estimates were also created using the LAPOP data. Again, no data were available for Santa Cruz, so $n_{skin\_brightness} = 23$.

2.4. Geographic variables

In line with Fuerst and Kirkegaard (2016a; 2016b), latitude and yearly mean temperature (in Celsius) of provincial capitals were used as climatic variables. We obtained these from the English language Wikipedia.

2.5. Analysis

The correlations between the demographic variables are shown in Table 13. The correlation between European BGA and SIRE was strong, though lower than found in the case of Brazil (r = .77; Fuerst & Kirkegaard, 2016b). The correlation between skin reflectance and European BGA was surprisingly low. Both Tierra del Fuego, for which we had no genetic data, and Santiago del Estero, which had substantial African ancestry (and so low brightness scores relative to other provinces, which are inhabited by mostly European and Amerindian hybrid populations), were major outliers. The relatively low correlations suggest that our ancestry indices are somewhat poor indicators of provincial admixture.

The primary variables of interest were correlated; results are shown in Table 14. For the LAPOP CA and S scores, we report aggregate-level factor analysis scores. As expected, the correlations between cognitive ability (CA), socioeconomic development (S), and the three indicators of European ancestry were moderate to strong and all positive, though this was not the case for HDI 2011 and European BGA. Figures 9 to 11 show the relationships between CA

(CA_One), S (S_Comb), and European BGA, with values weighted by the square root of the provincial population size.

Table 13. *Validation correlations for European ancestry, N = 23 provinces. BGA = biogeographic ancestry; SIRE = self-identified race/ethnicity. Unweighted correlations above the diagonal, correlations weighted by square root of population below.*

	Euro BGA	Euro SIRE	Skin reflectance
Euro BGA		0.69	0.22
Euro SIRE	0.67		0.48
Skin reflectance	0.33	0.48	

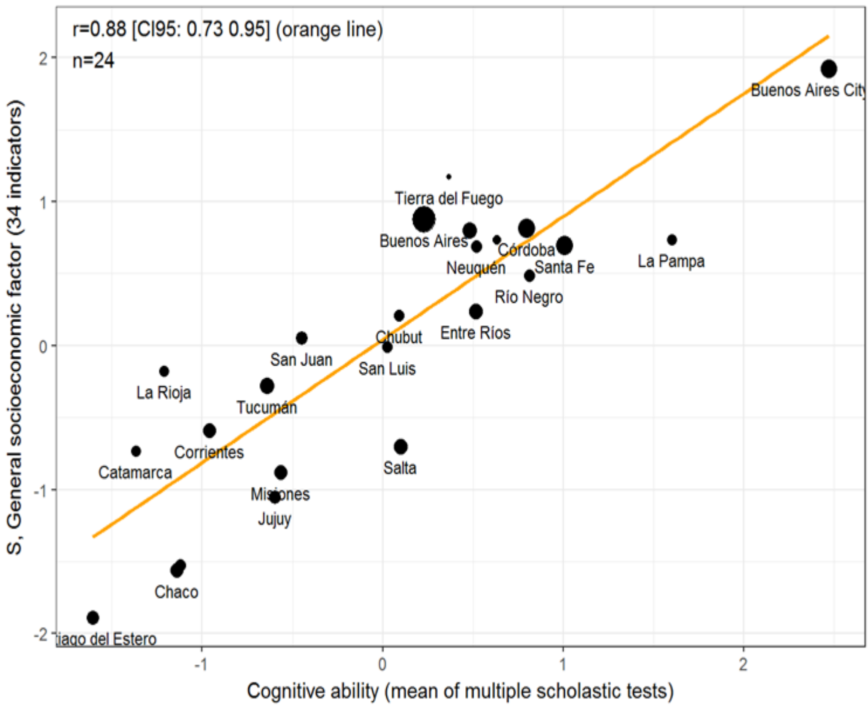

Figure 9. *Relationship between cognitive ability and general socio-economic factor.*

Table 14. *Correlations between the primary variables at the province level. CA, cognitive ability; S, socio-economic development; HDI, Human Development Index; BGI, biogeographic ancestry; SIRE, self-identified race/ethnicity. N = 23-24 provinces. Unweighted correlations below the diagonal, correlations weighted by square root of population above.*

	CA ONE	CA lapop	S knoema	S lapop	S comb	HDI 2011	HDI 1996	Euro BGA	Euro SIRE	Skin bright	Lat	Temp
CA ONE		0.65	0.85	0.75	0.88	0.87	0.81	0.48	0.47	0.69	0.47	-0.43
CA_lapop	0.60		0.65	0.77	0.75	0.65	0.54	0.44	0.53	0.57	0.50	-0.48
S_knoema	0.81	0.66		0.69	0.96	0.86	0.80	0.49	0.33	0.69	0.57	-0.50
S_lapop	0.75	0.75	0.66		0.86	0.71	0.55	0.56	0.67	0.71	0.65	-0.59
S_comb	0.87	0.75	0.95	0.86		0.87	0.77	0.54	0.48	0.76	0.61	-0.55
HDI2011	0.82	0.64	0.88	0.76	0.90		0.86	0.26	0.37	0.68	0.58	-0.54
HDI1996	0.76	0.56	0.83	0.63	0.82	0.87		0.42	0.32	0.51	0.46	-0.30
Euro_BGA	0.44	0.40	0.36	0.47	0.42	0.18	0.38		0.67	0.33	0.24	0.02
Euro_SIRE	0.50	0.52	0.31	0.69	0.50	0.36	0.33	0.69		0.48	0.36	-0.07
Skin_bright	0.65	0.55	0.62	0.68	0.71	0.68	0.54	0.22	0.48		0.51	-0.50
Latitude	0.48	0.53	0.58	0.69	0.64	0.70	0.57	0.12	0.38	0.56		-0.88
Temperature	-0.49	-0.54	-0.55	-0.65	-0.61	-0.71	-0.49	0.08	-0.19	-0.57	-0.93	

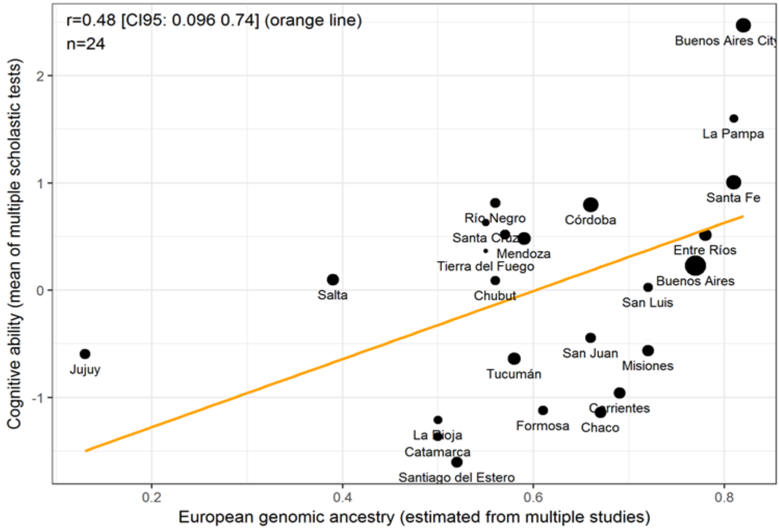

Figure 10. *Relationship between European biogeographic ancestry and cognitive ability.*

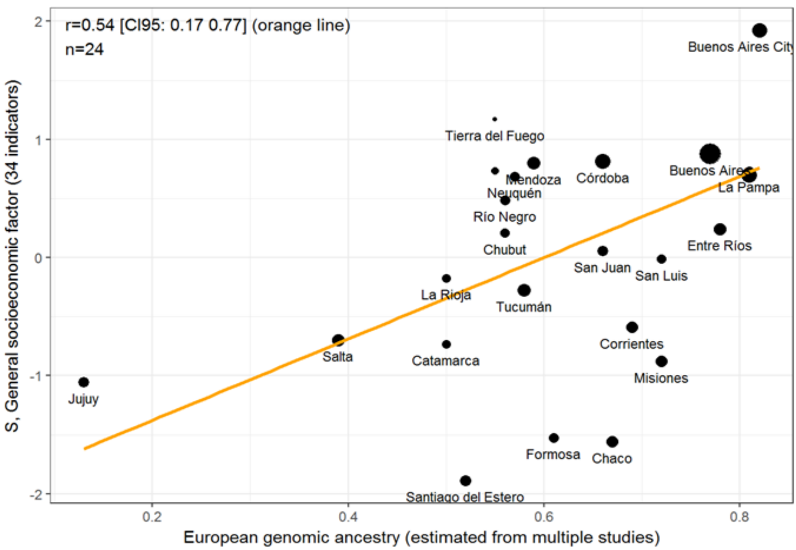

Figure 11. *Relationship between European biogeographic ancestry and general socio-economic factor.*

2.6. OLS regression

To examine how multiple predictors work together to predict CA and S, we used OLS regression as was done in previous studies (Fuerst & Kirkegaard, 2016a). For predicting CA, we included variables that plausibly could be taken to be causally prior to CA. We included only temperature and not latitude also because these predictors are highly correlated (r = -.93, weighted) which leads to very large standard errors when both variables are included, and because if latitude has any causal effect, it is likely mediated by a geoclimatic factor such as temperature. Results are shown in Table 15. The model was extremely overfitted as indicated by the negative cross-validated R^2. This is likely because of the small sample size and the use of weights. However, the results without weights were effectively the same (European = 0.48; temperature = -0.53). Despite this, European BGA did predict higher CA at above chance levels.

Table 15. *OLS full model results predicting cognitive ability. N = 24 provinces, weighted by square root of population size. R^2 = .43, R^2-cv = -2.3.*

Predictor	Beta	SE	CI lower	CI upper
European BGA	0.50	0.17	0.15	0.85
Temperature	-0.60	0.22	-1.06	-0.13

In the OLS regression for socioeconomic status (S), as above we entered the predictor variables that plausibly could be taken to be causally prior to S. Results are shown in Table 16. Despite the inclusion of cognitive ability (CA), which had a sizable beta, European biogeographic ancestry kept much of its zero-order validity. Temperature also had a strong negative beta.

Table 16. *OLS full model results for S (general socioeconomic factor). N= 24 provinces, weighted by square root of population size. R^2 = .84, R^2-CV = .40*

Predictor	Beta	SE	CI lower	CI upper
CA	0.65	0.12	0.41	0.90
European BGA	0.24	0.11	0.01	0.47
Temperature	-0.39	0.14	-0.68	-0.09

OLS regression tends to overfit models when too many predictors are used; this makes the results difficult to interpret (James et al., 2013). To overcome this, penalized regression methods such as LASSO regression were developed. These methods systematically shrink the betas of the predictors towards 0 and use within-sample cross-validation to select an optimal amount of shrinkage. The result is that the models contain fewer predictors and perform better in out-of-sample cross-validation tests. In line with some previous studies (Fuerst & Kirkegaard, 2016a; Kirkegaard, 2016a), we employed LASSO regression to estimate the best predictive model for the present dataset. As LASSO is a conservative method, null results cannot be taken as strong evidence of no validity unless sample sizes are large. LASSO regression for CA using the same predictors as above indicated that no predictor was reliable in the present dataset (500 runs). We repeated the above analyses for S. Table 17 shows the summary statistics from the LASSO regressions with S as the dependent variable.

Table 17. *LASSO regression betas summary statistics with S (general socioeconomic factor) as the dependent variable. CA, cognitive ability; BGA, biogeographic ancestry. N= 24 provinces, weighted by square root of population size. 500 runs.*

Statistic	CA	European BGA	Temperature
Mean	0.64	0.13	-0.24
Median	0.64	0.14	-0.25
SD	0.01	0.05	0.06
Mad	0.00	0.04	0.05
Fraction zero	0.00	0.01	0.00

We see that CA, European BGA, and temperature were useful predictors across (nearly) all runs. In this case, the LASSO essentially confirmed the findings from OLS. In the best model both CA and European had moderate to large positive betas while mean temperature was strongly negative.

2.7. Path analysis

Regression methods rest on the assumption that all predictors are causally independent and each causes or at least correlates with some cause of the outcome variable. This is rarely the case, especially in analyses of sociological data. Instead, one can posit explicit causal models using path analysis. Using this method, one can estimate both the direct and indirect effects of variables given a particular assumed model. We created a model incorporating temperature, European ancestry, CA, and S. While a climatological variable cannot cause someone to become more genetically European during their lifetime, it can lead to migrational patterns that do effect the average European ancestry of the provinces; for this reason, this variable was modeled as being causally prior to European BGA. The results are shown in Figure 12. The results show that the estimated relationships between European ancestry and CA/S hold when taking into account a more complicated causal network.

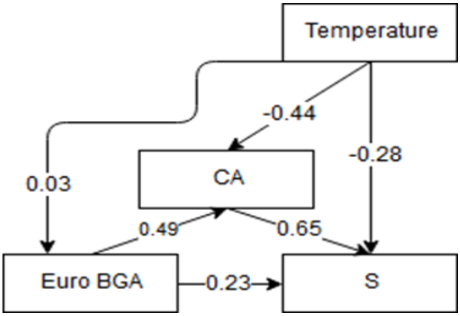

Figure 12. *Path model for European biogeographic ancestry (BGA), cognitive ability (CA), socioeconomic outcomes (S), and temperature. N = 24 provinces, weighted by square root of population size.*

2.8. Robustness check

Owing to the unexpectedly low provincial-level correlations between skin brightness and European ancestry (r = .33, weighted), we reran the path analysis with European SIRE and skin brightness as independents instead of European BGA. As expected, based on the correlations reported in Table 14, these indicators of ancestry showed robust associations with outcomes. That is, using different indicators of European ancestry did not substantially change the results. Figure 13 shows the path diagram with skin brightness as a predictor. Latitude is used instead of temperature as it is more plausibly causally related to brightness.

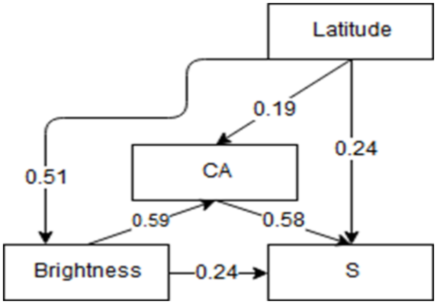

Figure 13. Path model for skin brightness, cognitive ability (CA), socioeconomic outcomes (S), and Latitude.

2.9. Taking into account all ancestries

In the previous sections, we looked at the association between European ancestry and outcomes. However, it is possible that combining two ancestries in multiple regression would improve the predictive power. Below, we present standardized betas for all three components. It is not possible to insert all three at once into a regression model, since the three ancestry values add up to 1 (perfect multicollinearity). As such, the betas for two at a time are presented. We have retained models with one predictor for comparison. We also report adjusted R as a measure of model fit. We caution that standardized betas, especially when weighted in multiple regression models, are not as easy to interpret as correlations. Results are shown in Tables 18 and 19. Generally, both African and Amerindian ancestry were associated with worse provincial outcomes. As the variance in African ancestry was low, those results need to be read with caution.

Table 18. *Multiple regression results for cognitive ability in Argentina. Each row represents one model. N = 24 provinces, weighted by square root of population size.*

African%	Amerindian%	European%	adj. R
-0.46			.38
	-0.40		.34
-0.42		0.49	.44
-0.31	-0.36	0.38	.48
		1.43	.48
	0.99		.48

Table 19. *Multiple regression results for socioeconomic development (S) in Argentina. Each row represents one model. N = 24 provinces, weighted by square root of population size.*

African%	Amerindian%	European%	adj. R
-0.38			.29
	-0.49		.43
		0.55	.50
-0.33	-0.46		.50
-0.19		0.48	.50
	0.62	1.14	.50

3. Municipal and district-level analyses

Argentina has multiple levels of administrative divisions: the first-level administrative divisions are the 24 provinces, below which are municipalities (n = 376), below which are departments and districts. We do not have extensive data for municipalities and districts, but with the LAPOP data it is possible to aggregate the scores to these levels. Since not all units at this level were sampled, we do not have complete coverage of the country. Furthermore, since the aggregate-level factor analyses seemed to generate somewhat more reliable scores, we used this scoring method. We included the two demographical variables which were calculated as before and added %White, which refers to the number of individuals who identified as "White". We weighted by the square root of the LAPOP unit sample size to take into account the uneven distribution of cases by administrative units. Results are shown in Tables 20 and 21.

Table 20. *Correlations for municipalities. S, socio-economic development; CA, cognitive ability; SIRE, self-identified race/ethnicity. N = 299, weighted by the square root of the LAPOP municipal sample size.*

	S	CA	Euro SIRE	White SIRE	Skin bright.
S	1.00				
CA	0.66	1.00			
Euro SIRE	0.33	0.20	1.00		
White SIRE	0.33	0.19	0.90	1.00	
Skin brightness	0.53	0.46	0.45	0.46	1.00

Table 21. *Correlations for districts. S, socio-economic development; CA, cognitive ability; SIRE, self-identified race/ethnicity. N = 437, weighted by the square root of the LAPOP district sample size.*

	S	CA	Euro SIRE	White SIRE	Skin bright.
S	1.00				
CA	0.52	1.00			
Euro SIRE	0.28	0.14	1.00		
White SIRE	0.32	0.13	0.89	1.00	
Skin brightness	0.43	0.28	0.44	0.48	1.00

In general, the lower-level results replicated those seen for provinces although with some decrease in effect sizes. This presumably is due to the decreased reliability of the measurements. Many of the samples were quite small, being based only on a few persons.

4. Discussion and Conclusion

As in previous studies, for aggregate-level data we found strong, positive relationships between cognitive ability (CA), the general factor of socioeconomic outcomes (S), and measures of European ancestry. The general pattern of results held for three different levels of aggregate analysis: provinces, municipalities, and districts (n = 24, n = 299, n = 437). The provincial relationships generally replicated in more complex analyses. However, while the mediation model for European BGA, CA, and S has generally been supported by previous analyses (Fuerst & Kirkegaard, 2016b), in the present study, much of the relationship between European BGA and S was not mediated by CA, despite the fair quality of cognitive data. As suggested below, this could be due to the relative crudeness of both the BGA and CA indicators and to the fact that our CA measure was of scholastic achievement, which is likely influenced by regional SES (thus, reverse causality). We will have to wait for better data to become available before investigating this issue.

We attempted to estimate European BGA at the municipal and district level using weighted SIREs. Our estimates only produced weak to moderate correlations with CA/S. Some of the weakness of the relationship was likely due to sampling error, and to measurement error for CA. Another possibility is that there is regional variation in the BGA proportions of the same SIRE groups, making SIRE an unreliable predictor of BGA at these levels of analysis. For instance, in Brazil we found that Whites (Branco) in Ceará (northeast) had a mean European BGA of 65% while Whites in Minas Gerais (southeast) had a mean of 89%. Unfortunately, the BGA data for Argentina were very sparse and so we were unable to examine this possibility.

At the individual level, we observed correlations between skin brightness, CA, and S, just as have been found in many previous studies (e.g., Lynn, 2002). These relationships were not due to SIRE-related social processing effects such as discrimination. We could not investigate whether the observed relationships involving skin brightness were due to racial phenotype-based social processes with this dataset. To determine this, one must have genomic ancestry data or use a between-sibling design (Dalliard, 2014). In general, the results are congruent with models that involve robust links between persons' ancestry and outcomes, and they are congruent with previous studies (Fuerst & Kirkegaard, 2016a, 2016b; Kirkegaard et al., 2017).

Interestingly, the correlation between European BGA and skin brightness was weak, despite both European BGA and skin brightness being robustly related to outcomes and despite our national-level analysis showing a strong relationship between European BGA and skin brightness (using the same methods and, for skin brightness, the same dataset). Why is not clear. The admixture estimates were based on unrepresentative samples; additionally, the estimates were likely somewhat unreliable owing to the number of molecular markers used. The LAPOP surveys similarly were not representative on the provincial level; moreover, interviewer assessed skin color is likely an unreliable measure of true color. As such, both indexes are likely very imperfect indexes of "true" provincial BGA; thus, it might be expected that the correlations between the ancestry indexes would only be modest.

One possible reason for why they were lower than expected is that in Argentina the source populations were predominately Europeans and Amerindians and thus the color contrast is not high, as in cases where Africans were a major source; in Argentina, African admixture is relatively low (unweighted African BGA mean $_{provincial}$: 6%, range $_{provincial}$ 1-17%). An alternative possibility is that one or the other measure is a poor index of actual provincial BGA. Skin brightness is substantially influenced by environmental factors (e.g., exposure to the sun), but also by genetic variance within continental races. As to this latter point, among indigenous Argentinians, as among Europeans, there is an evolutionary pigmentary cline which runs north to south (Chaplin, 2004). The non-random settlement of different European groups (e.g., south versus central/north) in Argentina would have increased the variance unrelated to continental level BGA.

Regardless, it is unlikely that this intra-continental BGA related variance could account for much of the provincial level associations between skin brightness and outcomes, which holds when including latitude, European BGA, and African BGA in path models (analyses not shown). On the provincial level, skin brightness is perhaps better thought to be indexing sociological factors (e.g., sun exposure related to occupation). It is possible that the same holds on the individual level, within a given province and given SIRE group. Given this concern, future studies should ideally use data with individual-level BGA, instead of relying on ancestry proxies such as SIRE and color.

5. Limitations

5.1. Individual analyses

First, the cognitive ability data was very limited, consisting of just a few items. 5/6 of the items concerned political knowledge (the last concerned geography), so there is likely a strong coloring of any general factor. In some analyses, we corrected for the estimated measurement error using the standard formula. However, such corrections are not easy to do for multivariate analyses and so were not done for these. Second, although the socioeconomic data were diverse, they were not entirely satisfactory. There was a lack of crime and health indicators and an abundance of property ownership indicators. This is likely to have colored the general factor to some degree, a claim supported by the moderate indicator sampling reliability observed. Third, we lacked individual level BGA data and so had to rely on skin brightness and SIRE as ancestry proxies.

5.2. Aggregate level analyses

First, there were only 24 units in the province-level analyses. Small data analyses tend to give unreliable results and important associations may go unnoticed due to lack of power/precision. For this reason, it is important to compile larger datasets, as done in Fuerst and Kirkegaard (2016a). Second, the quality of the estimates of BGA was moderate. There are likely to be fairly large estimation errors for multiple provinces that cause unknown errors in the results. Third, the study relied on scholastic test data instead of traditional cognitive ability (IQ test-based) data. Since scholastic ability is presumably more influenced by socioeconomic variables such as the quality of schooling (Branigan, McCallum & Freese, 2013), this limits the causal interpretations of the results, as there may be causation back and forth between scholastic ability and the socioeconomic outcomes. It is not possible to examine this issue with the present data. Fourth, due to the moderate quality of the predictors, in particular the European BGA variable, there is a high probability that some true validity of the causal variables was 'displaced' to other causal or non-causal predictors in the models. This is a direct statistical implication when one has correlated predictors with measurement error (Westfall & Yarkoni, 2016). It is likely that higher quality ancestry estimates will become available in the next couple of years, which can be used for an updated study. Fifth, the present dataset is cross-sectional. This makes it difficult to control for fixed-effects of provinces (e.g., latitude/average temperature) that are correlated with changing circumstances of the provinces (e.g., demographics). A better approach is to use a longitudinal design to control for fixed effects as done by Deryugina and Hsiang (2014) and Fulford et al (2016). Unfortunately, such datasets are hard to come by.

Supplementary material and acknowledgments

The R code, full datasets and high quality figures can be found at https://osf.io/etuy8/. Rpub can be found at http://rpubs.com/EmilOWK/ Argentina_admixture

We thank the Latin American Public Opinion Project (LAPOP) and its major supporters (the United States Agency for International Development, the Inter-American Development Bank, and Vanderbilt University) for making the data publicly available.

References

Branigan, A.R., McCallum, K.J. & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces* 92: 109-140.

Carl, N. (2016). IQ and socio-economic development across local authorities of the UK. *Intelligence* 55: 90-94.

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.

Chaplin, G. (2004). Geographic distribution of environmental factors influencing human skin coloration. *American Journal of Physical Anthropology* 125: 292-302.

Dalliard, M. (2014). The elusive X-factor: A critique of J. M. Kaplan's model of race and IQ. *Open Differential Psychology*.

DeMars, C. (2010). *Item Response Theory*. Oxford, New York: Oxford University Press.

Deryugina, T. & Hsiang, S. (2014). Does the environment still matter? Daily temperature and income in the United States. *NBER Working Paper* 20750.

Easterly, W. & Levine, R. (2016). The European origins of economic development. *Journal of Economic Growth* 21: 225-257.

Eppig, C., Fincher, C.L. & Thornhill, R. (2010). Parasite prevalence and the worldwide distribution of cognitive ability. *Proceedings of the Royal Society of London B: Biological Sciences* 277(1701): 3801-3808.

Fox, J. (2016). polycor: Polychoric and polyserial correlations (Version 0.7-9). Retrieved from https://cran.r-project.org/web/packages/polycor/index.html

Fuerst, J. & Kirkegaard, E.O.W. (2016a). Admixture in the Americas: Regional and national differences. *Mankind Quarterly* 56: 255-373.

Fuerst, J. & Kirkegaard, E.O.W. (2016b). The genealogy of differences in the Americas. *Mankind Quarterly* 56: 425-481.

Fulford, S.L., Petkov, I. & Schiantarelli, F. (2016). Does it matter where you came from? Ancestry composition and economic performance of U.S. counties, 1850-2010. *SSRN Scholarly Paper* 2608567. Retrieved from http://papers.ssrn.com/abstract=2608567

Herrnstein, R.J. & Murray, C.A. (1994). *The Bell Curve: Intelligence and Class Structure in American Life* New York: Free Press.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (eds.) (2013). *An introduction to statistical learning: with applications in R*. New York: Springer.

Jensen, A.R. (1998). *The g factor: the science of mental ability*. Westport, Conn.: Praeger.

Kievit, R., Frankenhuis, W.E., Waldorp, L. & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology* 4: 513.

Kirkegaard, E.O.W. (2014). The international general socioeconomic factor: Factor analyzing international rankings. *Open Differential Psychology.*

Kirkegaard, E.O.W. (2015a). IQ and socioeconomic development across regions of the UK: A reanalysis. *The Winnower*. Retrieved from https://thewinnower.com/papers/1419-iq-and-socioeconomic-development-across-regions-of-the-uk-a-reanalysis

Kirkegaard, E.O.W. (2015b, September 23). The general religious factor among Muslims: A multi-level factor analysis. Retrieved from http://emilkirkegaard.dk/en/?p=5485

Kirkegaard, E.O.W. (2016a). Inequality across US counties: An S factor analysis. *Open Quantitative Sociology & Political Science*.

Kirkegaard, E.O.W. (2016b). Some new methods for exploratory factor analysis of socioeconomic data. *Open Quantitative Sociology & Political Science*.

Kirkegaard, E.O.W., Wang, M.R. & Fuerst, J. (2017). Biogeographic ancestry and socioeconomic outcomes in the Americas: A meta-analysis of American studies. *Mankind Quarterly* 573: 398-427.

Knoema. (2013). Argentina regional dataset, October 2013. Retrieved from https://knoema.com/aboybld/argentina-regional-dataset-october-2013?region=1000250-binacional

La Griffe du Lion (2001, July). Pearbotham's Law on the persistence of achievement gaps. Retrieved August 16, 2015, from http://www.lagriffedulion.f2s.com/adverse.htm

León, F.R. & Burga-León, A. (2015). How geography influences complex cognitive ability. *Intelligence* 50: 221-227.

Lorenzo-Seva, U. & Ten Berge, J.M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology* 2(2): 57-64.

Lynn, R. (1979). The social ecology of intelligence in the British Isles. *British Journal of Social and Clinical Psychology* 18: 1-12. https://doi.org/10.1111/j.2044-8260.1979.tb00297.x

Lynn, R. (2002). Skin color and intelligence in African Americans. *Population and Environment* 23: 365-375.

Lynn, R. (2008). *The Global Bell Curve: Race, IQ, and Inequality Worldwide*. Augusta, GA: Washington Summit.

Lynn, R. (2015). *Race Differences in Intelligence*, revised edition. Augusta, GA: Washington Summit.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Programa de las Naciones Unidas para el Desarrollo. (2013). Informe nacional sobre desarrollo humano 2013. Argentina en un mundo incierto: Asegurar el desarrollo humano en el siglo XXI Buenos Aires. Programa de las Naciones Unidas para el Desarrollo. Retrieved from http://hdr.undp.org/sites/default/files/pnudindh2013.pdf

Putterman, L. & Weil, D.N. (2010). Post-1500 population flows and the long-run determinants of economic growth and inequality. *Quarterly Journal of Economics* 125: 1627-1682.

Rangel, M.A. (2014). Is parental love colorblind? Human capital accumulation within mixed families. *Review of Black Political Economy* 42: 57-86.

Revelle, W. (2015). psych: Procedures for Psychological, Psychometric, and Personality Research (Version 1.5.4). Retrieved from http://cran.r-project.org/web/packages/psych/index.html

Revelle, W. (2016). *An Introduction to Psychometric Theory with Applications in R*. Retrieved from http://www.personality-project.org/r/book/

Russo, M.G., Di Fabio Rocca, F., Doldán, P., Cardozo, D.G., Dejean, C.B., Seldes, V. & Avena, S.A. (2016). Evaluación del número mínimo de marcadores para estimar ancestría individual en una muestra de la población argentina. *Revista del Museo de Antropología* 9(1): 49-56.

Spolaore, E. & Wacziarg, R. (2013). How deep are the roots of economic development? *Journal of Economic Literature* 51: 325-369.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* 35: 401-426.

Telles, E.E. (2014). *Pigmentocracies: Ethnicity, Race, and Color in Latin America*, 1st edition. Chapel Hill, NC: University of North Carolina Press.

Templ, M., Alfons, A., Kowarik, A. & Prantner, B. (2015, February 19). VIM: Visualization and imputation of missing values. CRAN. Retrieved from http://cran.r-project.org/web/packages/VIM/index.html

Van de Vliert, E. (2013). Climato-economic habitats support patterns of human needs, stresses, and freedoms. *Behavioral and Brain Sciences* 36: 465-480.

Westfall, J. & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE* 11(3): e0152719.

Papers for the admixture estimates

Alfaro, E.L., Dipierri, J.E., Gutierrez, N.I. & Vullo, C.M. (2005). Genetic structure and admixture in urban populations of the Argentine North-West. *Annals of Human Biology* 32(6): 724-737.

Avena, S.A., Goicoechea, A.S., Bartomioli, M., Fernández, V., Cabrera, A., Dugoujon, J.M., ... & Carnese, F.R. (2007). Mestizaje en el sur de la región pampeana (Argentina). *Revista Argentina de Antropología Biológica* 9(2): 59-76.

Avena, S.A., Goicoechea, A.S., Rey, J., Dugoujon, J.M., Dejean, C.B. & Carnese, F.R. (2006). Mezcla génica en una muestra poblacional de la ciudad de Buenos Aires. *Medicina (Buenos Aires)* 66(2): 113-118.

Avena, S., Via, M., Ziv, E., Pérez-Stable, E.J., Gignoux, C.R., Dejean, C., ... & Beckman, K. (2012). Heterogeneity in genetic admixture across different regions of Argentina. *PLoS ONE* 7(4): e34695.

Avena, S.A., Parolín, M.L., Dejean, C.B., Ríos Part, M.C., Fabrykant, G., Goicoechea, A.S., ... & Carnese, F.R. (2009). Mezcla génica y linajes uniparentales en Comodoro Rivadavia (provincia de Chubut, Argentina). *Revista Argentina de Antropología Biológica* 9(1).

Avena, S.A., Parolin, M.L., Boquet, M., Dejean, C.B., Postillone, M.B., Alvarez Trentini, Y., ... & Carnese, F.R. (2010). Mezcla génica y linajes uniparentales en Esquel (Pcia. de Chubut): Su comparación con otras muestras poblacionales argentinas. *BAG. Journal of Basic and Applied Genetics* 21(1): 1-14.

Bobillo, C., Navoni, J.A., Olmos, V., Merini, L.J., Lepori, E.V. & Corach, D. (2014). Ethnic characterization of a population of children exposed to high doses of arsenic via drinking water and a possible correlation with metabolic processes. *International Journal of Molecular Epidemiology and Genetics* 5(1): 1-10.

Cano, H., Ginart, S., Caputo, M., Corach, D. & Sala, A. (2015). Analysis of the genetic structure of Santa Cruz province and itS comparison with the other Southern Patagonian provinces of Argentina. *Forensic Science International: Genetics Supplement Series* 5: e114-e115.

Corach, D., Lao, O., Bobillo, C., van Der Gaag, K., Zuniga, S., Vermeulen, M., ... & De Knijff, P. (2010). Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA. *Annals of Human Genetics* 74: 65-76.

García, A., Dermarchi, D.A., Tovo-Rodrigues, L., Pauro, M., Callegari-Jacques, S.M., Salzano, F.M. & Hutz, M.H. (2015). High interpopulation homogeneity in Central Argentina as assessed by ancestry informative markers (AIMs). *Genetics and Molecular Biology* 38: 324-331.

Godinho, N.M.O., Gontijo, C.C., Diniz, M.E.C.G., Falcão-Alencar, G., Dalton, G.C., Amorim, C.E.G., ... & Oliveira, S.F. (2008). Regional patterns of genetic admixture in South America. *Forensic Science International: Genetics Supplement Series* 1(1): 329-330.

Gómez-Pérez, L., Alfonso-Sánchez, M.A., Dipierri, J.E., Alfaro, E., García-Obregón, S., De Pancorbo, M.M., ... & Peña, J.A. (2011). Microevolutionary processes due to landscape features in the Province of Jujuy (Argentina). *American Journal of Human Biology* 23: 177-184.

Marino, M., Furfuro, S. & Corach, D. (2009). Genetic structure of Mendoza province population inferred from autosomal and Y-chromosome STRs analysis. *Forensic Science International: Genetics Supplement Series* 2(1): 433-434.

Martinez Marignac, V.L., Bertoni, B., Parra, E.J. & Bianchi, N.O. (2004). Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. *Human Biology* 76: 543-557.

Morales, J.O., Dipierri, J.E., Alfaro, E. & Bejarano, I.F. (2000). Distribution of the ABO system in the Argentine Northwest: Miscegenation and genetic diversity. *Interciencia* 25(9): 432-435.

Parolin, M.L., Avena, S.A., Fleischer, S., Pretell, M., Rocca, F.D.F., Rodríguez, D.A., ... & Manera, G. (2013). Análisis de la diversidad biológica y mestizaje en la ciudad de Puerto Madryn (Prov. de Chubut, Argentina). *Revista argentina de antropología biológica* 15(1): 61-75.

Parolin, M.L., Carreras-Torres, R., Sambuco, L.A., Jaureguiberry, S.M. & Iudica, C.E. (2014). Analysis of 15 autosomal STR loci from Mar del Plata and Bahia Blanca (Central Region of Argentina). *International Journal of Legal Medicine* 128: 457-459.

Resano, M., Esteban, E., González-Pérez, E., Vía, M., Athanasiadis, G., Avena, S., ... & Dejean, C. (2007). How many populations set foot through the Patagonian door? Genetic composition of the current population of Bahía Blanca (Argentina) based on data from 19 Alu polymorphisms. *American Journal of Human Biology* 19: 827-835.

Rocca, F.D.F., Albeza, M.V., Postillone, M.B., Acreche, N., Lafage, L., Parolín, M.L., ... & Avena, S. (2016). Historia poblacional Y analisis antropogenética de la Ciudad de Salta. *Andes* (27).

Rocco, F.D.F, de la Vega, D., Russo, G., Avena, S. (2013). El aporte africano al acervo génico de la población de Rosario, Provincia de Santa Fe. Conference: III Jornadas del Grupo de Estudios Afrolatinoamericanos. Buenos Aires: Ediciones del CCC Centro Cultural de la Cooperacion Floreal Gorini.

Seldin, M.F., Tian, C., Shigeta, R., Scherbarth, H.R., Silva, G., Belmont, J.W., ... & Alvarellos, A. (2007). Argentine population genetic structure: Large variance in Amerindian contribution. *American Journal of Physical Anthropology* 132: 455-462.

Sevini, F., Yao, D.Y., Lomartire, L., Barbieri, A., Vianello, D., Ferri, G., ... & Franceschi, C. (2013). Analysis of population substructure in two sympatric populations of Gran Chaco, Argentina. *PLoS ONE* 8(5): e64054.

Toscanini, U., Gusmão, L., Berardi, G., Gómez, A., Pereira, R. & Raimondi, E. (2011). Ancestry proportions in urban populations of Argentina. *Forensic Science International: Genetics Supplement Series* 3(1): e387-e388.

Trinks, J., Nishida, N., Hulaniuk, M.L., Caputo, M., Tsuchiura, T., Marciano, S., ... & Frías, S.E. (2017). Role of HLA-DP and HLA-DQ on the clearance of hepatitis B virus and the risk of chronic infection in a multiethnic population. *Liver Int.* 2017*;* 00*:*1-12*.* https://doi.org/10.1111/liv.13405*.*

Vieira-Machado, C.D., Carvalho, F.M., Silva, L.C., Santos, S.E., Martins, C., Poletta, F.A., ... & Orioli, I.M. (2016). Analysis of the genetic ancestry of patients with oral clefts from South American admixed populations. *European Journal of Oral Sciences* 124(4): 406-411.

Sex Differences on the Coloured Progressive Matrices in Sudan

Salaheldin Farah Attallah Bakhiet
King Saud University, Riyadh, Saudi Arabia

Nasser Sayed Gomaa Abdelrasheed
Dhofar University, Salalah, Sultanate of Oman

Helen Cheng
University College, London

Richard Lynn*
University of Ulster, UK

Yossry Ahmed Sayed Essa
King Saud University, Riyadh, Saudi Arabia

Tahanei Abdulrahman Muhammed Blahmar
King Saud University, Riyadh, Saudi Arabia

*Corresponding author: lynnr540@aol.com

Data are reported for a standardization of the Coloured Progressive Matrices in Sudan. Boys obtained a higher average IQ of 2.7 points and showed greater variability. The sample obtained a British-scaled IQ of 81.5.
Key words: Sudan, Coloured Progressive Matrices, Intelligence, Sex differences, Variability

One objective of this paper is to present data on sex differences on the Coloured Progressive Matrices for a sample of 6-9 year old children in Sudan. Sex differences on the Coloured Progressive Matrices for this age group had been given in a meta-analysis of 15 studies, most of them from Western countries, in which boys had an advantage of 0.21*d* equivalent to 3.2 IQ points (Lynn & Irwing, 2004). However, very little is known about sex differences in Sudan and other Arab countries, which are characterized by different sex roles and less effective school systems relative to Western countries. The second objective is to assess the performance level in Sudan in the international context, using norms from the British standardizations. British IQs for twelve studies of intelligence in Sudan have been given by Lynn and Vanhanen (2012) with estimates ranging between 64 and 86 and a median of 77.5.

Method and Results

The data have been published in Arabic as a standardization of Raven's Coloured Progressive Matrices (CPM) (Raven, Court & Raven, 1995) in Khartoum State, Sudan, in 2004 (Khatib, Mutwakkil & Hussain, 2006). The sample of 1683 school children (728 males, 955 females) was selected from public and private schools in the seven municipalities of Khartoum, the capital of Sudan, by stratified sampling. It was representative of all socio-economic levels. The subjects were tested in their schools by trained psychologists.

The results are shown in Table 1, separately for males and females of the four age groups of 6 through 9 years. This includes sample size N, mean raw scores and standard deviation (SD) on the CPM, the variability ratio (VR) calculated as male SD divided by female SD, the *d* values (average male score minus average female score divided by the average SD), and the t values as tests of the statistical significance of *d*. The last three columns compare the Sudanese scores with those in the British standardizations of 1982 and 2007 (Raven, Court & Raven, 1995; Raven, 2008). The lower percentile scores in the 2007 than the 1982 British standardization indicate strong Flynn effect gains for this age group in Britain during this 25-year period. The last column gives the Sudanese IQ according to British norms in 2004. It is calculated from IQs according to the 1982 and 2007 norm tables: IQ2004 = (IQ1982 x 3) + IQ2007 x 22) / 25. The calculation assumes constant yearly gains for British children between 1982 and 2007.

Discussion

There are four points of interest in the results. First, in all four age groups the boys obtained higher average scores than the girls. The average male advantage is 0.178 *d* and is equivalent to 2.7 IQ points. This male advantage is closely similar to the 0.21*d* (3.2 IQ points) male advantage given in the meta-analysis of 15 studies by Lynn and Irwing (2004) in which boys had an advantage of 0.21*d* equivalent to 3.2 IQ points. Second, boys showed greater variability than girls. This confirms numerous studies reporting that "Males have a slightly but consistently wider distribution than females at both ends of the range" (Deary, Penke & Johnson, 2010).

Table 1. *Sex differences on the Coloured Progressive Matrices in Sudan. ** p<.01.*

Age	Sex	N	Score ± SD	VR	*d*	*t*	British percentile 1982	British percentile 2007	British IQ 2004
6.5	M	92	13.5±5.4	1.26	.206	1.53	21	7	78.5
	F	167	12.5±4.3				14	5	76
7.5	M	167	14.2±4.6	0.92	.042	0.43	11	2	70.5
	F	293	14.0±5.0				10	2	70
8.5	M	286	16.1±5.8	1.09	.144	1.81	16	0.9	66.5
	F	366	15.3±5.3				9	0.5	63.5
9.5	M	183	17.6±7.3	1.24	.318	2.80**	8	0.8	65
	F	129	15.5±5.9				4	0.3	60

Third, the mean IQ of the Sudanese sample relative to the British standard at the time of testing (2004) is 70.1 for boys and 67.4 for girls. This can be averaged to 68.75 rounded to 69.This figure needs to be corrected because the test was administered to groups rather than individually and the test has been standardized in Britain for individual administration. A previous study has shown that in an Arab sample of this age group that took the test individually and in groups, the group administration gave a score 12.5 IQ points lower than individual administration (Bakhiet & Lynn, 2014). Thus, we propose that 12.5 needs to be added to the 69 to give 81.5 as the best estimate of the British-scaled IQ of the present sample. This is reasonably consistent with the median IQ of 77.5 in Sudan of previous studies noted in the introduction.

Fourth, it will be noted that the British IQs of the Sudanese children declined with age. The 6 year olds obtained IQs of 78.5 (boys) and 76 (girls), while the 9 year olds obtained IQs of 65 (boys) and 60 (girls). This decline is consistent with Jensen's (1977) cumulative deficit hypothesis stating that an adverse environment has a depressing effect on intelligence that increases with age. Jensen had demonstrated such an effect for black children in the rural South of the United States.

Acknowledgement

The authors extend their appreciation to the Deanship of Scientific Research, King Saud University for funding this work through the International Research Group Project RG-1438-007.

References

Bakhiet, S.F.A. & Lynn, R. (2014). A study of the IQ in Palestine. *Intelligence* 47: 10-11.

Deary, I. Penke, L. & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience* 11: 201-211.

Jensen, A.R. (1977). Cumulative deficit in IQ of blacks in the rural South. *Developmental Psychology* 13: 1841-1891.

Khatib, M., Mutwakkil, M. & Hussain, A. (2006). Tagneen I'khtibar al-masfofat al-mutatabia' al-mulawwan litalameez al-halaqa al-o'la bimarhalat al-asas biwilayat al-khartoum (Standardization of the Coloured Progressive Matrices for children in Khartoum State). Khartoum: Sharikat Matabi' al-Sudan lil-Omlah al-Mahdodah.

Lynn, R. & Irwing, P. (2004). Sex differences on the Progressive Matrices: A meta-analysis. *Intelligence* 32: 481-498.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Raven, J. (2008). *Coloured Progressive Matrices and Crichton Vocabulary Scale Manual.* London: Pearson.

Raven, J.C., Court, J.H. & Raven, J. (1995). *Coloured Progressive Matrices*. Oxford, UK: Oxford Psychologists Press.

MANKIND QUARTERLY 2017 57:4 585-589

A Study of the Intelligence of South Sudanese Refugee Children

Habab Abdlahiy Osman
Tabuk University, Tabuk, Saudi Arabia

Salaheldin Farah Attallah Bakhiet*
King Saud University, Riyadh, Saudi Arabia

Asma Sirageldin Fathelrahman Ahmed, Nourallah Ahmed Ganem Alnaier
University of Khartoum, Khartoum, Sudan

Richard Lynn
University of Ulster, UK

*Corresponding author: slh9999@yahoo.com

Data are reported for the intelligence of a sample of 2,990 South Sudanese refugee children aged 7 to 18 years tested with the Standard Progressive Matrices in 2016. The average British-scaled IQ of this sample is estimated to be at or below 55.
Key words: South Sudan, Standard Progressive Matrices, Intelligence, IQ tests, Refugees.

The southern part of Sudan seceded to become the independent nation of South Sudan in July, 2011. The principal reason for this secession was that the people of the southern part of Sudan are ethnically different from those of the center and the north. A genetic analysis by Cavalli-Sforza, Menozzi and Piazza (1994, Figure 3.9.1, p. 181) showed that the South Sudanese are principally a Nilotic people, one of the four sub-races of Negroids (Baker, 1974, p. 329), while the people of central and northern Sudan are a mixed race population consisting principally of North African Caucasoids (K.C.B., 1960, p. 512-3).

An early study of the intelligence of the Nilotic people in South Sudan had been carried out in 1954 by Fahmy (1964). He administered four tests to a sample of 291 school children aged 7-16 years, whom he described as "one of the primitive Nilotic Negro tribes" (p. 164) inhabiting the west bank of the White Nile and belonging to the Shilluk people. The four tests and the mean American-scaled IQs obtained by the sample were 73.5 on the Goddard Formboard, 76.5 on the Porteus mazes, 94.4 on the Alexander Passalong, and 53.4 on the Goodenough Draw-a-Man (DAM). The average IQ on the four tests is 74.5. The norms for these American tests were collected in the 1920s and 1930s and the Sudanese IQs do not take account of the increases in intelligence of approximately 3 IQ points a decade that have taken place in the United States (Flynn, 1984). Lynn and Vanhanen (2012) make an adjustment for the increases in intelligence to reduce the IQ of the Sudanese sample to 69 but this is not a satisfactory study because of the differences in the IQs obtained from the four tests, the long time interval between the date of the American norms and the collection of the data in Sudan, and the more than sixty years since the study was carried out. We report here a recent study of the intelligence of the South Sudanese.

Method

The sample consisted of 2990 children (1584 males, 1405 females) from South Sudan, age 7-18, at schools in two refugee camps that were established in Sudan for those adversely affected by the wars in South Sudan. They attended schools supervised by the Sudanese Red Crescent Society and funded by UNHCR, WFP and UNICEF. The first camp was in Alsalam and held 2524 children from the Shilluk tribe, and the second was in Algablain and held 466 children from the Nuer tribe. The children and teachers completed research participation consent forms. The children were tested in 2016 with the Standard Progressive Matrices (SPM) test (Raven, Raven & Court, 2000) without time limits by four trained psychologists. The SPM is a non-verbal test of visual comprehension and reasoning. It is widely used as a "culture-reduced" test of cognitive development in cross-cultural research because it does not require knowledge that is taught explicitly in school.

Results

Table 1 gives the numbers, SPM scores and British-scaled IQs of the sample. Because of strong British Flynn effects for children up to about age 12 (though not for teenagers), IQs were calculated separately from the norm tables of the British 1979 and 2007 standardizations. In the case of the 2007 standardization of the SPM+ (an updated, more difficult form of the test), the British-scaled IQs were calculated by converting the Sudanese SPM raw scores to SPM+ scores using Table SPM3 in Raven, Raven and Court (2000), and calculating the British IQs from Table A1 in Raven (2008). The mean British IQ of the twelve age groups cannot be determined precisely because most of the scores are outside of the norming range of the British standardizations. The 2007 standardization produces estimated IQs below 55 for all ages older than 8 years. Results are similar when the IQs are calculated from the British 1979 norms, which again produce IQs below the lower limit of the norming range.

Table 1. *Standard Progressive Matrices raw scores and IQs according to British norms from the test standardizations in 1979 and 2007.*

Age	N	Score ± SD	British-scaled IQ	
			1979	2007
7	187	10.8 ± 6.3	75	63
8	159	10.3 ± 4.7	≤70	59
9	279	11.4 ± 4.9	≤70	<55
10	271	10.9 ± 5.5	<70	<55
11	303	12.5 ± 7.3	<70	<55
12	273	13.8 ± 7.7	<70	<55
13	336	14.3 ± 7.8	<70	<55
14	295	17.7 ± 10.8	<70	<55
15	263	16.8 ± 11.1	<70	<55
16	202	20.5 ± 12.8		<55
17	222	22.2 ± 12.7		<55
18	182	24.1 ± 13.5		<55

Discussion

The main result of this study is an extreme floor effect that was observed through the entire age range. The SPM test consists of 60 items, each with either 6 (sets A and B) or 8 answer choices (sets C, D and E). Therefore random guessing alone will earn an average raw score of 8.5 items guessed correctly. The results of this South Sudanese sample show that up to the age of 10 years, very few children were able to solve any of the visual puzzles of this test, including the very first items that require simple visual matching rather than abstract reasoning.

Above age 10 we do see the expected rise in scores with increasing age, but the results are outside the norming range of the British 2007 standardization. For example, for 18-year-olds in the 2007 standardization of the SPM+, the corresponding SPM score at the 50th percentile of the distribution is 49. This defines an IQ of 100. At the 0.4th percentile, which marks an IQ of 60, British 18-year-olds have a raw score of 32. Therefore a precise IQ cannot be calculated from the South Sudanese raw score of 24.1. A more meaningful comparison is that the mean score of 18-year-old South Sudanese is at the 25th percentile of British 7-year-olds in 2007 (IQ 90), or at the 60th percentile of British 7-year-olds in 1979 (IQ 105). At age 15, which is the oldest age included in the British standardization of 1979, the British median (IQ 100) was 47 raw score points and the 5th percentile (IQ 75) was 33 raw score points, compared with the South Sudanese mean of 16.8 points.

The British-scaled IQ of less than 55 is one of the lowest recorded for any sample from sub-Saharan Africa. Lynn (2015, p. 59) gives 143 studies of the IQ of Negroid samples for which the median is 71. The lowest IQ in this compilation is 60 for studies in Gambia and Tanzania. The average IQ of the present South Sudanese sample is comparable to that of some minor peoples from sub-Saharan Africa. It is similar to the IQ of 55 for the Bushmen and the IQ of 57 for the Pygmies given in Lynn (2015, pp. 122, 126).

The low IQs of the present subjects are likely attributable in part to the deprivations, interrupted life and education that they experienced before the move to the refugee camps and while they were in the camps. A number of them were orphans whose parents had been killed in the war. Most had been at schools before they were admitted to the camps, but quality of schooling has never been high in South Sudanese villages, and most of them had suffered some degree of malnutrition. Their food in the camps consisted of three meals a day mainly of lentils, beans, bread and corn but with little meat and possibly a shortage of vitamins and minerals. It is inevitable that rudimentary education, nutritional deficiencies and disruptive early life experiences have contributed to poor cognitive development. Improving these conditions remains a task for the future.

Acknowledgement

The authors are thankful to the Deanship of Scientific Research, and the Department of Education and Psychology at Tabuk University for funding this research.

References

Baker, J.R. (1974). *Race*. Oxford, UK: Oxford University Press.

Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1994). *The History and Geography of Human Genes.* Princeton, NJ: Princeton University Press.

Fahmy, M. (1964). Initial exploring of the intelligence of Shilluk children. *Vita Humana* 7: 164-177.

Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin* 95: 29-51.

K.C.B. (1960). Sudan. *Encyclopedia Britannica* 21: 504-513.

Lynn, R. (2015). *Race Differences in Intelligence: An Evolutionary Analysis*, 2nd edition. Augusta, GA: Washington Summit Publishers.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Raven, J. (2008). *Standard Progressive Matrices-Plus version and Mill Hill Vocabulary Scale Manual.* London: Pearson.

Raven, J., Raven, J.C. & Court, J.H. (2000). *Standard Progressive Matrices*. Oxford: Oxford Psychologists Press.

MANKIND QUARTERLY 2017 57:4 590-631

Lineage Population: A Concept Needed by an Observer of Nature?

John Fuerst*
Ulster Institute for Social Research, London, UK

*Address for correspondence: j122177@hotmail.com

The genealogy-based classificatory programs of Kant and Darwin are briefly discussed for context. It is detailed how in biology there is no unambiguous term to reference infraspecific-level descent-based divisions. The term *lineage population* is introduced and defined for analytic purposes as one of a set of inter-fertile divisions of organisms into which members are arranged by propinquity of descent. It is argued that the lineage population concept avoids the ambiguities associated with related biological and anthropological concepts and polysemes such as *population*, *ethnicity*, and *race*. Other terms and concepts, such as *form*, *cline*, *cluster*, *geographic population*, *breeding population*, *genetic population*, *breed*, *species*, *subspecies*, *ancestry*, *geographic ancestry*, *biogeographic ancestry*, *ancestral population*, *ancestry population*, *natural division*, and *population lineage*, are discussed in relation to this concept. It is concluded that the lineage population concept is a useful analytic tool which describes, in line with the Kantian/Darwinian perspective, an interesting class of biological variation.
Key Words: Systematics, Kant, Darwin, Lineage population, Population, Ethnicity, Race, Population lineage

Descent was first proposed as an organizing principle for natural history during the 18th century,[1] but only during the 19th century, with the acceptance of

[1] Natural history (lt. *historia naturalis*), in the 18th and 19th century sense, referred to the study of nature and natural objects. The focus encompassed, among other domains, zoology, botany, and biological anthropology. While the applied biological sciences, such as animal husbandry and agriculture, were frequently discussed and made reference to

evolutionary theory, did it become widely adopted. Despite general acceptance of phyletic-based[2] classification and analysis, nomenclature has remained confused and ambiguous, especially on the infraspecific level. While the term *race* was previously used to denote descent-based groupings either on the specific, subspecific, or infrasubspecific level, a number of biologists and anthropologists have come to feel that the word is problematic and that more precise terms are called for (for example: Wagner et al., 2016). Unfortunately, alternative terms and concepts are presently either ambiguous or nonspecific, respectively. Heeding the call for precision, the neutral term, *lineage population*, is introduced and defined to correspond with Darwinian infraspecific and specific "communities of descent." This concept is discussed in relation to other biological and anthropological terms and concepts.

1. Systematics and the Kantian/Darwinian principle

Over two hundred years ago, Immanuel Kant (1777, 1788) argued that the guiding principle for natural history, insofar as one wished to understand patterns of variation, should be genealogical relationship. He contrasted *scholastic* with *natural* divisions and stated that "a natural division is based upon identifying lines of descent that classify animals according to reproductive relationship" (1777). The Kantian principle for natural history was summarized in 1796 by Christoph Girtanner, who wrote that a new Kantian "system of classification for the animal kingdom into classes, orders, species, races, variations, and varieties [developed] according to relationships of generation, must be taken up." The principle was likewise critiqued by naturalist Georg Forster who argued, on the basis of Linnaean systematics, that it was impossible to trace infraspecific lines of descent. Forster (1786) notes: "[The situation] must be different in natural history, if, in this [field] we are, as Kant claims, concerned only with generative origination and descent. Natural history in this sense might, however, possibly be only a science for gods and not for human beings." Kant's descent-based principle for natural history led directly to his conception of race, which was seen as necessary "for an observer of nature… from the viewpoint of natural history" (1788). *Races* were one of the manifold of possibilities which could potentially characterize biological variation understood from a genealogical perspective; conceptually, they were "progenies *classificae*," lines of descent or deviate forms that "are still

---

in order to explain natural historian theories and concepts, these fields were generally not considered to be part of natural history proper.

[2] This is a disputed term. Meant here is Mayr's Haeckelian sense of "pertaining to lines of descent." See: Mayr & Bock (2002).

so distinct and persistent that they justify a class distinction" (1788); in practice, they were infraspecific groups of organisms which exhibited hereditary characters with a uniformity sufficient to allow for a descent-based classification.[3]

Kant is of primary interest because he was the first to clearly recognize and to forcefully articulate the view that a "natural" system for natural history would be a descent-based one (Doron 2012). His conceptualization had a lasting impact because both his perspective on natural history and his race concept were adopted by others, such as the influential naturalist Johann Blumenbach (Lenoir 1980). Linnaean systematics having been overturned,[4] the descent-based organizing principle is now foundational to biology, including the subfields of taxonomy, conservation biology, evolutionary biology, and population genetics. As such, groups of organisms – on the infraspecific, specific, and supraspecific level – are conceptually related to one another in terms of genealogical affinity since propinquity of descent, "the only known cause of the similarity of organic beings, is the bond, hidden as it is by various degrees of modification" (Darwin 1859, 413-414). The rationale for this organizing principle was articulated by economist and philosopher William Jevons in *Principles of Science* (1874, 680-719):

> It is true that in the biological sciences there would be one arrangement of plants or animals which would be conspicuously instructive, and in a certain sense natural, if it could be attained, and it is that after which naturalists have been in reality striving for nearly two centuries, namely, that *arrangement which would display the genealogical descent of every form from the original life germ*. Those morphological resemblances upon which the classification of living beings is almost always based are inherited resemblances, and it is

[3] Conceptually, for Kant, *race* simultaneously described the lineage diagnosing characters which were transmitted along lines of descent; the lines of descent themselves; and the classes of organisms which were delineated by these characters, such as skin color (Kant, 1785) and possibly skeletal form (Kant, 1788). In practice, though, races inevitably referred to classes of organisms.

[4] Linnaeus' views evolved but his early ones were most influential. Accordingly, the members of Linnaean species, representing originally distinct lineages, were understood as being united in descent; and both supraspecific and infraspecific groupings of organisms (Linnaean classes, orders, and varieties) were only resemblance-based. Linnaeus conducted hybridization experiments over his career and discovered that certain of his species and other higher order groupings were interfertile, which led him to conclude that many of what he originally called species were not originally distinct lineages, which were expected to be intersterile.

evident that descendants will usually resemble their parents and each other in a great many points…. There is no reason to suppose that the same kind of natural classification which is best in biology will apply also in mineralogy, in chemistry, or in astronomy.

While Kant's views were indirectly influential,[5] it was Charles Darwin who established the primacy of descent for classifying organisms and understanding the relationships amongst them. Darwin's position was quite similar to that later expressed by Jevons (1874). For Darwin, a descent-based classification was seen as both a natural one and one in which systematics[6] should be grounded, because commonality in descent explained most organismic similarity and because arranging by descent had the effect of grouping organisms by overall phenetic similarity, which is what early systematicists, from Linnaeus on,[7] strove for. In this descent-based system, *race* found a place in natural history as a description of specific and infraspecific communities of descent. Commenting to Thomas Huxley in 1867 on a *natural system* and a classification of human races, Darwin noted:

I knew, of course, of the Cuvierian view of Classification, but I think that most naturalists look for something further, & search for "the natural system", – "for the plan on which the Creator has worked" etc. etc. – It is this further element which I believe to be simply genealogical….

Grant all races of man descended from one race; grant that all structure [i.e., physical features] of each race of man were perfectly known – grant that a perfect table of descent of each race was perfectly known – grant all this, &

---

[5] Kant's definition of *race*, for example, was cited in Blumenbach's widely read *A Manual of the Elements of Natural History* (1825, 15-16).

[6] Here, Mayr and Ashlock's (1991, 431) broad conception of systematics is adopted: "The science dealing with the diversity of organisms." Taxonomy, the "theory and practice of classifying organisms," is nested within this. By this term usage, a "breed" is a systematic classification or unit but not a taxonomic category/rank.

[7] Bowler (1989, 65) succinctly notes: "It must be supposed that in the Creator's plan *every* relationship has a meaning; thus a truly natural system of classification will have to take into account all the characteristics of every species. Linnaeus believed that such a natural system was the goal of his work, although at the beginning of his career, he was overwhelmed by the immense amount of information that would have to be processed in order to set up a natural system. As a preliminary set, he decided to establish an 'artificial system' that would classify according to resemblance in a *single* characteristic."

then do you not think that most would prefer as the best classification, a genealogical one, even if it did occasionally put one race not quite so near to another, as it would have stood, if allocated by structure alone. Generally, we may safely presume, that the resemblance of races & their pedigrees would go together. (Darwin, 1903, letter 204)

1.1. Semiotic ambiguity of race, past and present

Prior to the 18th century, the French term *race*, which in ordinary usage primarily meant lineage[8], was not infrequently used to refer to humanity or to human divisions understood in terms of lineage. Thus, for example, historian Sir Paul Rycaut (1668) could comprehensibly discuss groups of the Ottoman Empire in a decidedly genealogical fashion[9], and many of the groups described as *races* by Rycaut, such as Tartars or Turks, correspond with *races* discussed by 18th century natural historians, such as Buffon (1777). Yet, prior to the 18th century, such *races* found no place in a coherent biological systematics for the simple reasons that a modern concept of biological species – which race, as a natural scientific concept, has always been defined in relation to – was missing and that humans were not treated as part of natural history. For the former one had to wait for John Ray (Wilkins 2009, 65-67) and for the latter Carl Linnaeus (Smith 2015, 136-138). Contrary to what is sometimes said or suggested (for example: Lieberman, 1968), Linnaeus himself did not introduce *race* – either the term or a unique concept which might reasonably be called this – into natural history. On the contrary, the systematic categories which he popularized, those of *variety* and *species*, left no epistemic space in natural history for *race* (at least one not redundant with *species*) (Doron 2012; see footnote 11). It was both because it

---

[8] For example, in *A Dictionarie of the French and English Tongues* compiled by Randle Cotgrave in 1611, *race* is defined as: "a linnage, familie, kindred, house, bloud; litter, brood; sort, kind…"; in Pierre Richelet's (1757) *Nouveau dictionnaire françois*, *race* is defined as: "lignee. Extraction. Desendans. Famille. Sorte de gens. Espece"; in Noël and Chapsal's (1832) *Nouveau dictionnaire de la langue française*, *race* is defined as "(radix, gen, icis) extraction, origine; lignee; tousceux qui viennent de meme famille."

[9] Rycaut (1668) notes, for example, that: "They that are of this Race never dare vaunt of their Pedigree, it is a contumaciousness and almost Treason to name it; nor have I learned that there is any Family amongst the Turks of this Line, of any account or esteem, but one alone, who is called Ibrahim Hun Ogleri, or the Off-spring of Sultan Ibrahim, their Father being a Son of the Grand Signior's Sister, and married to a Sultana (and are said to be of the Race of the Tartars) so that proceeding by the Women's side only, the less notice is taken of their Blood…"

became obvious that there was something to explain – the constancy of so-called constant varieties – and it was often unclear if these propagating forms represented infraspecific as opposed to specific lineages that the terminology of *race* was adopted.

Darwin (1871) and others recognized that the term *race* lent itself to confusion as, in the context of natural history, it could refer to species, subspecies, hereditary varieties, and domestic breeds – all understood as lineages and/or forms which propagate themselves across generations to form lineages. While the term's pedigree seems to have been well appreciated, there was, from the start, disagreement regarding how restrictively it should be employed. Thus contra Kant (1777), who delimited the term *race* to describe infraspecific divisions,[10] Forster (1786) rejected the Kantian principle in favor of Linnaean systematics with its species versus inconstant variety distinction,[11] a distinction which made no room for Kant's concept of race. Forster argued that the term race should be used in an undetermined way to describe lineages, whether they constituted, as he saw the alternatives, Linnaean species or Linnaean varieties. Understood in the latter sense, races were "defined by changeable, accidental characteristic features" such that "one variety can change into another" and had no real place in natural history (1786). Forster writes:

---

[10] Though, in his *Lectures on Anthropology*, Kant frequently refers to the "human race," meaning the human species; so across works he is not consistent on the matter. But, for an exclusively infraspecific definition, see also Duchesne (1766, 18-26) and, by way of Kant, Blumenbach (1825, 15-16).

[11] Müller-Wille (2007) notes: "At the core of Linnaeus's reform stood his distinction between species and variety which was thoroughly based on his theory of generation. In distinguishing between species, Linnaeus advised his fellow naturalists, one should rely exclusively on "constant" characters – that is, not on characters that varied with external conditions like climate or nutrition, but on characters that reproduced in offspring under various external circumstances." Briefly, in Linnaean systematics, species were *defined* in terms of separate origins (but see footnote 3) and *diagnosed* by their ability to reproduce their form across generations. Varieties, in contrast, were *defined* as environmentally induced deviations from the species form and they were *diagnosed* by their inconstancy across different environments. It was soon realized that some peculiarities within what seemed to be the same species line of descent were constant (across environments), which led to the oxymoronic term *constant varieties* and to lumper-splitter debates as to whether various groups, given the relative constancy of their forms, were *varieties* or *species*.

I have deliberately made use of the word *variety* in the previous discussion, but [I have] at the same time given it to be understood that I consider it synonymous with [the word] *race*….
We have borrowed [the term] from the French; it seems very closely related to [the words] *racine* and *radix* and signifies descent in general, though in an indeterminate way. For one talks in French of the race of Caesar [in] the same [way] as of the races of horses and dogs, irrespective of the first origin, but, nevertheless, as it seems, always with tacit subordination under the concept of a species…. [The word] should mean nothing more than a mass of men whose common formation is distinctive and sufficiently at variance with their neighbors [such that they] could not be *immediately* derived from them. [They are] a lineage whose derivation is unknown, and consequently, one which we cannot easily count under one of the commonly accepted human varieties because we lack knowledge of the intermediary link. Thus, the Papuans and the other black inhabitants of the islands of the South Sea related to them are called a different *race* [distinguishable] from the light brown people of Malaysian descent that can be found in the same region. [This, however, is only to say] that [they are] *a people of peculiar character and unknown descent.*

Forster was hardly the only one who had something to say about term usage. French naturalist Jean Baptiste Bory de Saint-Vincent and Marcellin (1827, 63) criticized monogenists for trying to evade the conclusion that different human groups were different species "by referring to 'races,' probably not remembering that the word race, synonymous with lineage, is most usually used in speaking of domestic animals…" For Bory de Saint-Vincent, as for Forster, *race* did not have a real, non-redundant place in (Linnaean) systematics – thus groups were either species or Linnaean varieties.[12] Approaching the matter from the other side of the debate, natural historian Armand de Quatrefages (1861, 311) excoriated the American polygenist school for "assimilating" the term *race* and for failing to draw

[12] Bory de Saint Vincent (1827, 70) notes: "It would be necessary to prove that whites and negroes differ due to climate in which they live, that the lineage of negroes and whites have changed, without crossbreeding from white to black, or from black to white, after having been carried from south to north or from north to south." Here, he is employing the species/inconstant variety dichotomy to defend the claim of the plurality of human species.

the proper distinctions between the *race* and *species* concepts. De Quatrefages writes:

Let us first notice in MM. Nott and Gliddon the complete assimilation of the two words race and species. – In Europe, all botanists, all zoologists, from Linnaeus to de Candolle, from Buffon to Cuvier, and Geoffroy Saint-Hilaire, have employed them to designate very different things. If some have designated race by the expression hereditary variety, this difference in words does not in any way affect ideas.
The distinction which exists all facts considered is always translated into language. Yet it is this distinction that the American school seems to forget entirely here. For her, there are no more races or varieties in nature; there are only species.

De Quatrefages' historiography is problematic as, for example, Buffon and Cuvier called both specific and infraspecific lineages *races*, but that is another matter. What is relevant is that race concepts were disputed and that the term was employed differently by different researchers, situations which led to confusion and ambiguity. At issue was not only whether race, properly understood, referred exclusively to infraspecific or inclusively to both infraspecific and specific lineages. As Bory de Saint-Vincent's comment suggests (cited above), some thought that the proper denomination of the term was to refer to domestic breeds. Thus, Deniker (1906, 4-8) questioned whether one can properly speak of human races in a technical zoological sense, since according to him, the term typically describes "domestic animals living under artificial conditions." He opted to use race in "a very broad sense, different from that given to it in zoology and zootechnics." Paul Broca (1864) noted another concern: that certain polygenists were using the term in a "general and misleading sense" to refer to collections of (Linnaean) species. Broca criticized these researchers on account of their uncritical and self-defeating adoption of the "ambiguous term race" (1864, 14) to refer to, in addition to species and infraspecific lineages, polyphyletic groups. He noted:

[These other polygenists] continue also very often to use the term *race* to designate the ensemble of all individuals of each group, adopting this by a sort of transaction of language of those whose system they reject; and thus they speak of the white or Caucasian race… as if, for instance, the brown Celts and the fair-haired Germans had descended from the same primitive stock. This contradiction has given a handle to monogenists; for if climate and mode of life can cause a German to become a Celt, there is no reason why, under certain influences, a Celt might not become a Berber, a Berber a Foulah, a Foulah a Negro, and a Negro an Australian.

Of course, when discussing the "human species" or the "human race" polygenists (for example: Virey (1837, 102-104) would also have employed terms in misleading senses, since humanity, so conceptualized, would constitute a polyphyletic group equivalent to a Linnaean supraspecific division. With the acceptance of evolutionary theory, the classic ontological distinction between specific, infraspecific, and supraspecific classes vanished and with it some of these semantic concerns. It was no longer that, as Linnaeus put it, "Nature makes species and genus, culture makes varieties, art and nature makes classes and orders" (Fundamenta Botanica 1737, qtd. in Stamos 2005) and so one could treat infraspecific races, species, and closely allied species as roughly the same type of thing. Before the acceptance of evolution, the term *race* allowed one to write ambiguously of population lineages without committing oneself to a position as to the origin of these. After, one could do the same, but without eliding an important ontological distinction. This generic usage, while having utility, as noted by Gray (1879, 320-321),[13] has contributed nontrivially to the cloud of confusion surrounding the term. For example, it led William Boyd (1950, 186-187), hardly an opponent of race concepts, to conclude, based on his mid-20th century etymological analysis of the word *race*, that the "implication of the separate origins for different races came inevitably…." The sense of *race* Boyd is speaking of, though, is none other than Forster's *species*, and Forster and other polygenists,[14] as suggested above, recognized that the *race* notion undercut the force of their argument for separate origins. It was part of, as Paul Broca (1864, 10) put it, the "language of those whose system" the polygenists rejected. That is, contra Boyd, the generic notion of *race* was employed by monogenists in their explanation of how groups could transmit their character differences across generations – could

[13] Gray (1879, 320-321) notes: "Wherefore, since we hardly need the term race in the restricted sense of seed-propagating variety, it is sometimes convenient to use it in the manner proposed by Bentham … as the common designation of any group or collection of individuals whose characters are continued through successive generations, whether permanent variety, subspecies, species, or group consisting of very similar species, the term not implying any decision of this question. If this use of the term prevails, S*ubspecies*, will probably take its place as the designation of the highest grade of variety."

[14] The term polygenist is used in the original sense of those who argued that different human groups represented separate creations or autochthonic lineages and thus separate species.

be forms which propagated themselves across generations – without being separate Linnaean species.[15]

Between the middle of the 18th century to the second half of the 19th century, it was frequently debated whether human and other lineage-based divisions were *races* qua Linnaean species or *races* qua varieties.[16] These debates were substantive in that interlocutors agreed on a common, non-conventional definition of species – originally distinct lineages – but disagreed when it came to what constituted evidence for separate origins. As Wallace (1864) points out, monogenists argued that continuance in variation, a lack of homogeneity, interfertility between groups, and a difficulty in delineating groups evidenced that the various races were constant varieties (for example: Prichard 1836). Polygenists, on the other hand, rejected these diagnostic criteria and emphasized that popularized by Linnaeus.[17] During the 20th century and continuing into the

---

[15] See, for example, Pierre Louis Maupertuis's account of inherited differences between human races in *Venus physique* (1745), an account which was picked up by the widely influential Comte de Buffon. See, also, Duchesne's (1766, 18-19) discussion of why the concept of race was needed so to prevent one from being forced to rank human (constant) varieties as separate species: "It often happens that instead of the word Race, Species is employed; which obliges species to be given the name of genera; thus, instead of saying that in the human species, there are several races, that of white men, that of black men, etc.; whites and blacks are regarded as distinct species, which make up the human race. I confess that this way of expressing oneself would be more in accordance with the etymologies which I have just related… but this usage is actually not that of scientists…."

[16] In botany and zoology there were "lumper-splitter" or "variety-species" debates parallel to the anthropological monogenist-polygenist ones (see: Stamos 2007), though the term *race* was less commonly used in these discussions to denote lineages in the most general sense.

[17] Often missed in discussions is that the polygenist understanding of species was very minimalistic; species were only expected to exhibit at least one hereditary character difference (Broca & Blake 1864; Forster 1786; see also Wallace's (1864) characterization of the position). Thus, for example, Forster took the constancy of pigmentation differences between Europeans and Black Africans as evidence that the races were separate Linnaean *species*, as opposed to separate Linnaean *varieties*. He notes: "Is the Negro a *variety* or a *species* in the human genus? If [the ruling in this matter] depends upon proving the descent of all varieties from an original, common parental couple, which cannot be demonstrated without indisputable historical evidence, there will be no definite solution… If, on the other hand, we are satisfied by the Linnaean ruling, [that] a variety differs from a species simply through the inconstancy of its characteristic features, then a little provisional investigation is still required [to find out] to what extent this definition fits

21st, anthropologists began debating whether the human divisions previously called *races* really qualified as *races*.[18]

Whereas the 18th to early 19th century species/variety debates were substantive, these were and are mostly semantic, turning on the users' specification of the term race. During this time, many proponents of the "no-races" position narrowly constructed the meaning of *race* – so that the word implied markedly discontinuous populations, separate homogeneous populations, populations between which there were extensive genetic differences, divisions between which there had been extensive reproductive barriers, etc.[19] – to mean something closer to the polygenists' Linnaean species and, on the basis of these understandings, deemed that race, the construct, had either no biological validity or no applicability to humanity. Others, deeming that *race* implied populations between which there were phyletic discontinuities[20] significant enough to enable

---

the various human lines of descent." Virey (1837, 19-35) makes roughly the same argument: "In whatever light we consider Negroes, we cannot deny that they present characteristics of a race distinct from the white. This truth, grounded upon incontestable facts of anatomy, is universally acknowledged. Now, in natural history, that which distinguishes a *species* from a *race,* is the permanency of the characteristic features, notwithstanding contrary influences of climate, food, or other external agents: whereas, races are but varied modifications of a sole and primordial *species*… [W]hy do [negroes] not remain white in cold countries, and when kept from the light? If the blackness of their skin was produced by a cause entirely occasional and external why should it be hereditary in all countries, and the same in all generations?...Everything serves to prove that negroes form, not only a *race*, but undoubtable a *distinct* species, from the beginning of the world."

18 See Lieberman (1968), Lieberman and Reynolds (1978), and Littlefield et al. (1982) for an outline of the history of the human "race-exists" and "no-race" positions debate.

19 See Lieberman, Kirk and Corcoran's (2000) Table 2. The authors note: "Lieberman and Kirk's survey of 1999 asked respondents about their support or rejection of biological race and found that among those who rejected the race concept 79 percent supported analyzing variation in terms of clines rather than races, 78 percent rejected the idea of homogeneous populations, 80 percent supported more variation within so called races than among them and, for 88 percent, gene flow invalidated labeling distinct races (Table 2)." It is not made clear why *races* came to be seen as entailing these characteristics. Based on the characterization of the "19th century idea of race" which the authors seem to accept, poor historiography seems to have played a part.

20 Littlefield et al. (1982), for example, note: "More recently, the race concept has been attacked as invalid because populations of humans separated by significant reproductive

the objective partitioning of the species, concluded that the human divisions previously called races did not count as bona fide natural historian *races*.[21]

In defense of this position it was claimed that human races do not exist because human groups “grade one into another” (Brace 1964) and because “[b]oundaries between what have been called 'races' are completely arbitrary, depending primarily upon the wishes of the classifier" (Brace 1964; Lieberman 1968). On this point, Lieberman and Reynolds (1978) tell us that “[t]he issue is not whether there is factual proof of hereditary variations” but whether character variation is distributed “in such a discordant pattern that identifiable boundaries cannot be established...” Yet, originally, the central issue was, in fact, the nature and status of hereditary variation, that is, of seed-propagating forms, and not what would now be called population structure. Regarding the latter topic, as noted above, gradation, especially in the human genus, was well recognized and was the basis for one common argument for why divisions were races *in the infraspecific sense*, that is, were hereditary varieties – a point which Brace (1964) and Lieberman (1968) tacitly acknowledge with their remarkable reference to none other than Blumenbach as the originator of this argument.[22] At the same time, theoretical paradigms changed, and proponents of the position that (human) races exist correspondingly updated their concepts.[23] The result of the conceptual

barriers and/or exhibiting concordant combinations of variable physical traits cannot be shown to exist.”

21 Some, such as Ashley Montagu (1942), granted that, of course, human races exist in a zoological and genetic sense, but argued against supposed popular conceptions, such as, for example, race as the “prime determiner of all the important traits of body and soul, of character and personality, of human beings and nations” (pp. 9).

22 Similarly, Smith (2015, 256) notes: “For a reifier of race, we see here, Blumenbach certainly cedes quite a bit to what we today would call the constructionist camp.” The difference is that Brace (1964) and Lieberman (1968) – along with many contemporaneous anthropologists and philosophers of biology – see the “constructivism,” well-recognized by early “reifiers of race,” involved in infraspecific racial classifications, especially human ones, as an argument against these classifications.

23 Some, for example Hochman (2013), have argued that proponents of the position that human races exist have watered down the concept of race; however, it’s hard to reconcile this position with typical 18th and 19th century usages, in which the term was broadly employed to refer to both species and constant varieties and as the latter was conceptualized in a very “weak naturalistic” sense; for example, Gray (1879, 320) defines race, in the narrow technical sense, as “a variety which is perpetuated with considerable certainty by sexual propagation.”

revisions is that there are multiple disputed meanings of the term (Lieberman et al. 2004).

Unfortunately, contemporaneous biological definitions, often being less than transparent, have done little to help resolve the semiotic confusion. While contemporaneous dictionaries of biology, genetics, and zoology frequently define the term (for example, Martin & Hine 2008; Allaby 2013[24]), the definitions are often opaque and not obviously tethered to the historic sense of lineage. Compare these definitions with Gould (1894) and Beeton (1871),[25] where the understanding of lineage, stock, or propagating form is more obvious.

By these definitions, human and non-human races obviously exist – in the way that other groupings such as morphs, forms, and demes do (i.e., the concepts have applicability) – but it is not clear to what extent the term refers to a concept or conceptual network reasonably similar to what it once generally did. For example, in the 8th edition of *A Dictionary of Genetics*, King, Mulligan, and Stansfield (2013) define *subspecies* and *race* in the following manner:

subspecies 1. a taxonomically recognized subdivision of a species. **2.** geographically and/or ecologically defined subdivisions of a species with distinctive characteristics. *See* race. (p. 456)

race a phenotypically and/or geographically distinctive subspecific group, composed of individuals inhabiting a defined geographical and/or ecological region, and possessing characteristic phenotypic gene frequencies that distinguish it from other such extension groups. *Homo sapiens* can be subdivided into five races on the basis of geographic origin. (p. 391)

---

[24] Martin and Hine (2008, 500) define race as: "(in biology) A category used in the classification of organisms that consists of a group of individuals within a species that are geographically, ecologically, physiologically, or chromosomally distinct from other members of the species… (in anthropology) A distinct human type possessing several characteristics that are genetically inherited." Allaby (2013, 478) defines race as: "An interbreeding group of individuals all of whom are genetically distinct from the members of other such groups of the same species. Usually these groups are geographically isolated from one another, so there are barriers to gene flow."

[25] Gould (1894, 1231) defines race as: "In biology, a genealogic, ethnic, or tribal stock; a breed or variety of plants or animals made permanent by constant transmission of its characters through the offspring...." Beeton (1871, 239) defines race as: "In Bot., those permanent varieties of species which can be propagated by seed."

King, Mulligan, and Stansfield's *subspecies* are formally recognized *races*; they are taxa assigned to the taxonomic category or rank immediately below that of species.[26] But what their *races* are and how they are precisely delineated is unclear. According to their definition, *races* are "populations"[27] which differ in hereditary phenotypic characters. When it comes to humans, *races* are said to be divisions based on "geographic origin," not contemporaneous geographic location; thus *races* – at least human ones – are not strictly geographic populations. Imaginably, if pressed on the matter, the authors would make the same types of clarifications that O'Brien and Mayr (1991) and Mayr (2002) have.

According to O'Brien and Mayr (1991), who attempt to clarify the meaning of *subspecies*, the members of a subspecies "share a unique geographical range or habitat, a group of phylogenetically concordant phenotypic characters, and a unique natural history relative to other subdivisions of the species." The sharing of phyletic-based characters and a unique natural history implies genealogical affinity between members – and thus that the concept aligns with that of *natural divisions* in the Darwinian sense and finds a meaningful place in a phyletic-based taxonomy. Regarding human races, Ernst Mayr (2002) tells us:

> A human race consists of the descendants of a once-isolated geographic population primarily adapted for the environmental conditions of their original home country. But, as is illustrated by the success of Europeans and Africans and Asians in all parts of the world, any race is capable of living anywhere.

Accordingly to Mayr (2002), the human species was recently comprised of populations, ones relatively geographically isolated for a number of generations, which "agree in most characteristics with the geographic races of animals." Presently, he tells us, human races are the sum of the – presumably, relatively purebred – descendants of these isolated populations. Since members of human races can live anywhere, for Mayr, propinquity of descent seems to stand as the classifying criterion for *race* in the most general sense. Perhaps King, Mulligan,

---

[26] A taxon (pl. taxa) is a particular group of organisms that is given a formal name (e.g., *Hippopotamus amphibius*) and assigned to a definite taxonomic category (e.g., species). A taxonomic category is a rank in a hierarchal ordering of life.

[27] A *population* is defined as "a local (geographically defined) group of conspecific organisms sharing a common gene pool; also called a *deme*" (p. 370). A *deme* is defined as "a geographically localized population within a species" (p. 120).

and Stansfield (2013) mean this when they speak of “geographic origins.” Thus a member of their “Asian” or “Mongoloid” race whose family relocated to Brazil and then back to Japan would not be double-raced merely on account of their double geographic origin. However, the authors’ meaning is unclear.

1.2. Semiotic ambiguity, subspecies, and a concept needed

In *Origin of Species* (1859, 51), Darwin recognized that there exists a continuum running from individual variation to species and that hereditary varieties fell along this. He tells us:

> Certainly no clear line of demarcation has as yet been drawn between species and sub-species... or, again, between sub-species and well-marked varieties, or between lesser varieties and individual differences. These differences blend into each other by an insensible series; and a series impresses the mind with the idea of an actual passage.

Darwin recognized that both adaptation and isolation were important factors in what he recognized as the process of speciation. Between species and individual variants, Darwin recognized subspecies, “forms which possess many of the characteristics of true species, but which hardly deserve so high a rank” (*Descent of Man,* 1871, 227) and (constant) varieties, for which a “community of descent is almost universally implied, though it can rarely be proved” (*Origin of Species*, 1859, 44). Darwin situated these communities of descent in his genealogy-based classificatory system. As he notes:

> In confirmation of this view, let us glance at the classification of varieties, which are believed or known to have descended from one species. These are grouped under species, with sub-varieties under varieties; and with our domestic productions, several other grades of difference are requisite, as we have seen with pigeons. The origin of the existence of groups subordinate to groups, is the same with varieties as with species, namely, closeness of descent with various degrees of modification.... In classing varieties, I apprehend if we had a real pedigree, a genealogical classification would be universally preferred; and it has been attempted by some authors. For we might feel sure, whether there had been more or less modification, the principle of inheritance would keep the forms together which were allied in the greatest number of points.... If it could be proved that the Hottentot had descended from the Negro, I think he would be classed under the Negro group, however much he might differ in colour and other important characters from negroes. (Origin of Species, 1859, 426-427)

This extended passage is interesting for several reasons. First, Darwin has in mind a descent-based hierarchical scheme of classifications of organisms which extends well below the level of subspecies, a system of classification which blurs insensibly into individual differences, the raw material for the production of varieties. Second, he understands subspecies to be lineage-based divisions which "come very near to" species. While similar to contemporary taxonomic understandings, insofar as "subspecies" expresses a "taxonomically significant" magnitude of divergence, this conception is markedly different from original ones, as the term *subspecies* was originally employed to describe infraspecific constant varieties in general. For example, defining the term in *Hannoversches Magazin* (1784), Swiss botanist Jakob Ehrhart notes:

> Halbarten, Scheinarten, Subspecies.
> In this way I term plants which agree in essentials almost completely with each other, and are often so similar to each other that an inexperienced person has trouble in separating them, and about which one can conjecture, not without reason, that they have formally had a common mother, notwithstanding that they now always reproduce their like from seed. They are in a word, Varieties constantes, or an intermediate between *species* and *Spielarten* [inconstant varieties]. They are separated from species in that they differ from one another in small particulars of little importance; and they differ from *Spielarten* in that they reproduce themselves unchanging by seed and always beget their like. (qtd. in Chater, Brummitt & Ehrhart, 1966)

According to Ehrhart, subspecies were varieties which passed on their often minor character differences through their seed, which is to say *constant varieties* or the same type of entities which Comte de Buffon referred to as the *races* of a species.[28] It was only in the mid-19th to early 20th century – and so largely for convenience sake – that *subspecies* came to refer to constant varieties which were more like species in terms of the degree of differentiation, that is, to "the highest grade of variety" (Gray 1879, 321). During the 20th century, *race*, for the most part, was delimited such to refer to infraspecific descent-based divisions in general, thus including both subspecific and infrasubspecific divisions but excluding specific ones. As such, in line with Mayr and Ashlock (1991), King,

---

[28] For example, Buffon (1778, 252), equating infraspecific races with constant varieties, notes: "Races in each species of animal are only constant varieties that are perpetuated by generation, whereas in vegetable species there are no races, no varieties fairly constant to be perpetuated by reproduction."

Mulligan, and Stansfield (2013) define *subspecies* to mean formally recognized races – ones assigned a trinomen and so treated as taxa.

What is germane is that during the 19th century, the Linnaean conception of variety – and species – was rejected, and that through the mid-19th to the early 20th century, *subspecies* replaced *variety* as the designation for the taxa immediately below species in both botany and zoology. In zoology, unlike botany, no infrasubspecific category came to be recognized. Nonetheless, during the same period, zoological *races* below the level of subspecies were recognized, not formally as part of, to use Kant's phrase, a "systematic description or nature" but as biological classifications, or systematic units, used to understand biological variation (Simpson, 1969;[29] Wright, 1978[30]). Today, Darwin's infrasubspecific varieties would often be, ambiguously, called *populations*. Since *population* can mean geographic population and breeding populations, and since neither of such type of divisions necessarily need to be communities of descent, an important concept, for understanding biological variation, is presently missing. Given this situation, in conjunctions with the difficulty in precisely delimiting specific, subspecific, and infrasubspecific boundaries (Darwin 1859; 1871; Mayden 1997; Mayr 2000), a general concept to describe lineage-based variations, or communities of descent, would be useful.

Regarding such a general concept, Beatty (1985, 278) points out that Darwin treated species and hereditary varieties as the same kind of entity. He notes and asks:

They were referring, Darwin believed, to chunks with[in] the genealogical nexus of life. They did not refer to *one kind* of chunk with their species names and to *another kind* of chunk with their variety names.

---

[29] Simpson (1969, 101-102) notes: "In biology the term "race" is not a systematic category but it is used for any local infraspecific breeding group that is conveniently distinguished for purposes of a given study. The term "subspecies" is used in the technical sense for a race considered sufficiently distinct, uniform, and widespread to merit a Latin name… However, there are many smaller groups which cannot be assigned in a really valid way to one of these major divisions, and it is equally good biology to apply the term race loosely to such groups of all sizes and degrees of distinction."

[30] Wright (1978, 439) notes: "There is also no question, however, that populations that have long inhabited separated parts of the world should, in general, be considered to be of different subspecies by the usual criterion that most individuals of such populations can be allocated correctly by inspection… It is, however, customary to use the term race rather than subspecies for the major subdivisions of the human species *as well as for minor ones."* [Italics added.]

This raises the question why Darwin did not at least define a joint "species-variety" category in genealogical terms?

Part of the answer to Beatty's question, one which Beatty himself fails to note, I would suggest, is that Darwin employed the term *race* to refer to this "species-variety" chunk of the genealogical nexus as a *systematic unit*. That is, he was not without a term to describe the concept; he just did not designate a corresponding *taxonomic rank*.

2. The concept of lineage population

One might try to rehabilitate the term *race* – to bring it back in line with its historic meaning of lineage – and then specify a meaning to give it a definite place in contemporaneous phyletic-based systematics. However, the definition of the term is so disputed and riddled with confusion that it is desirable to clarify a needed concept and refer to it by a neutral term. But what is to be described? From the genealogy-based perspective, we notice a type of biological variation in want of a clear concept to pick it out and a minimally ambiguous term to reference this concept. After all, what term describes so-called *Europeans*, what the International HapMap Project refers to as "CEU," living in both Tuscany and Utah, or so-called *East Asians* in the same places? These groups are evidently not spatial populations, and they need not be either breeding populations or species. These are groups of organisms, not characters, whether discontinuous or clinal,[31] singular or multiple. One might call them *forms* (Štrkalj 2000), but this term is imprecise, not distinguishing between individual variants and populations, and it seems to neglect an important aspect regarding how these groups are delineated. These types of groups will be called *lineage populations*. They are defined as: *intra*-fertile divisions of organisms, where membership is determined by shared ancestry as revealed by ancestry-informative markers, and whose distinctiveness has been maintained by endogamy. In this context, *population* is used in a biostatistical sense – to mean, simply, a group of organisms with some features in common – and is used interchangeably with *division*, *class*, and *variable-type*. If typological thinking simply involves thinking in terms of individual characteristics

[31] The meaning of the term *cline* shifted over the 20th century. The term originally signified a character gradient – a distribution of traits (Huxley 1939). Referring to this meaning, Simpson (1961, 179) notes: "A cline is an arrangement of characters, not of organisms or of populations." Yet, *cline* has come to be used, often ambiguously, to also describe a population continuum – which is a type of distribution of organisms.

which allow one to classify individuals into groups, this is unabashedly a variable-typological conception as defined by Weiss and Lambert (2011).

In biology, the term is often used to describe cell clonal lineages, but it is also used, at times, to characterize the type of groups being discussed.[32] Since we are looking for a neutral term, the infrequency of term use, with respect to organismic groups, is desirable. The concept of lineage population can be contrasted with those of *phenon*, *variant*, or *morph*, which describe intrapopulational variants (Mayr and Ashlock 1991). The distinction would be similar to that made by Kant (1777) with respect to races and individual varieties such as blondes and brunettes or by Comte de Buffon (1777, 555-565) with respect to races and individual varieties such as albinos. Individual variants do not pick out separate lineages and are thus not separate lineage populations. The lineage population concept further differs from the concept called *form*, a neutral term used to describe individuals, phena (phenotypes within a population), or taxa (Mayr and Ashlock 1991). The concepts differ in that lineage populations, as defined here, do not necessarily refer to conspicuously distinct groups – for they could be molecularly and physiologically distinct[33] – and that forms do not necessarily refer to lineage-based divisions.

More centrally, the concept of lineage population can be juxtaposed with those of spatial and breeding population, both of which also describe groups of organisms. In biology, the term *population* has a number of different meanings (Waples & Gaggiotti 2006). As Schaefer (2006) notes, a "unified population concept remains elusive." Frequently, the term is used to refer to a *geographic population* defined as "a group of individuals of the same species occupying a particular space at a particular time" (qtd. in Waples and Gaggiotti 2006). For example, when, in *Descent of Man*, Darwin discusses the "population of the United States" (pp. 131) and the "immense mongrel population of Negroes and Portuguese" (pp. 225) in Brazil, he is referring to the geographic sense of the term. Alternatively, *population* is used to mean "a group of interbreeding individuals," or a *breeding population*. *Populations* can also refer to "collections of individuals that share some biological attributes" (Waples and Gaggiotti 2006), or what are here called *biostatistical populations*.

---

[32] For example: Schrode et al. (2012).

[33] Martin and Hine (2008, 196) define a form as "any distinct variant within a species." Perhaps for them and others one can have molecular and physiological cluster-forms, in which case lineage populations would be a sort of forms.

Finally, *population* is now frequently used to mean what Aulchenko (2010) refers to as a *genetic population* in the retrospective sense. These are biostatistical populations which are delineated in terms of ancestry-indexing characters. As Aulchenko (2010) notes, the term *genetic population* can also mean breeding population. For example, Robinson (2013, 111) tells us: "Conveniently, a genetic population may be defined as a group of individuals freely interbreeding, the limits of which may be either prescribed or implied." This ambiguity leads Aulchenko (2010) to distinguish between *genetic populations* in the retrospective and prospective senses. For some, the term *genetic population* seems to imply jointly being a breeding population and a lineage population. For example, Verspoor, Stradmeyer, and Nielsen (2008, 484) define a *genetic population* as "a group of sexually reproducing individuals and their relatives, within which mating is more or less random but among which interbreeding is constrained, so that they constitute a distinct gene pool." That different genetic populations have distinctly different gene pools owing to linebreeding entails that these are lineage populations. Likewise, Mukherjee (2007, 6) defines a *genetic population* as "a reproductively isolated group of individuals of the same species that share in a common descent or a common gene pool…." Here again genetic populations are lineage populations. However, these definitions seem to stipulate that these populations are, at the same time, breeding populations. As defined in this paper, lineage populations do not *necessarily* have to be breeding ones. To take an extreme case, there could be two reproductively isolated groups between which barriers suddenly, completely collapsed. As a result, one would immediately have one breeding population and two distinct lineage populations. This breeding/lineage population distinction allows for the possibility of *hybrids,* as recently exogamously mating individuals maintain their original lineage population status, despite collectively forming one breeding population; this allows their offspring to be of mixed lineage population.

While it is not clear how common this understanding of *population* is, Fujimura et al. (2010) find that, based on their interviews, genome-wide association study (GWAS) researchers "specify populations based on concepts of genetic ancestry." If this is a general practice, it would be consistent with Jiménez-Tejada et al.'s (2008) finding that unlike in ecology textbooks, in genetic ones, *population* is rarely defined in the geographic sense. If the term *population* is frequently understood to mean lineage population, one might argue that appending "lineage," for the sake of clarity, is unnecessary. Yet, that explicitness is needed is indicated by the fact that it is not readily obvious to Fujimura et al. (2010) and other authors, including the present one, when *population* means what is here called *lineage population*. That said, by *lineage population* something similar to Aulchenko's (2010) retrospective *genetic population* is meant, although genetic relatedness might be operationalized differently.[34]

To further clarify, this concept does not imply phyletic discontinuity between classes; thus, lineage populations, like 18th and 19th century infraspecific races,[35] can be picked out of a continuum formed by primary or secondary intergradation. Darwin recognized that lineages exist in a phyletic continuum. Regarding an evolutionary continuum brought about by the resurrection of forms long past, Darwin (1862, 330-331) notes:

> It is due to their absence, and to the consequent wide gaps in the series, *that we are enabled to divide the existing species into definable groups*. If there had been no extinction, there would still have been great lines, or branches, of special development ... but ancient and intermediate forms, very different probably from their present descendants, *would have rendered it utterly impossible to separate by distinct characters*... one great body from [an]other. [Emphasis added.]

His point is that phyletic disjunctions enable, or authorize, one to divide by distinct character biodiversity into definable evolutionary groups. In absence of disjunctions, the lineages still exist; they are merely hidden under a continuum of variation. In this case, distinct lineage populations cannot be objectively picked out any more than distinct colors can from the electromagnetic spectrum. In the same way, given similar circumstances, different geographic and breeding populations cannot be objectively delineated. Nor, under any circumstance, can

---

[34] Aulchenko's (2010) definition suggests using pairwise genetic distance to arrange members; however, one could, alternatively, delineate groups based on distance from population centroids.

[35] As Wallace (1864) noted, a lack of discontinuity was one of the common arguments made in defense of the position that various human groups *were* infraspecific as opposed to specific *races*: "In favour of the unity of mankind it is argued that there are no races without transitions to others; that every race exhibits within itself variations of colour, of hair, of feature, and of form, to such a degree as to bridge over to a large extent the gap that separates it from other races. It is asserted that no race is homogeneous; that there is a tendency to vary; that climate, food, and habits produce and render permanent physical peculiarities, which, though slight in the limited periods allowed to our observation, would, in the long ages during which the human race has existed, have sufficed to produce all the differences that now appear. It is further asserted that the advocates of the opposite theory do not agree among themselves; that some would make three, some five, some fifty or a hundred and fifty species of man…."

different temporal species. Nor can subspecific taxa which fall along a population continuum.[36] All of these divisions are nonetheless recognized.

In the wild, there is an obvious relation between geographic, breeding, and lineage populations, as for relatively sessile and range-restricted organisms, geographic populations of the same species will often be breeding populations. Since these breeding populations are often closed as opposed to *open* or "freely exposed to gene flow" (Mayr & Ashlock 1991, 142), over time they become lineage populations. Owing to the conceptual covariance, distinctions between types of populations are frequently not made. Nevertheless, these specifications need not coincide, as is the case with many zoo animals around the world. Thus, members of the great northern hippopotamus lineage population might find themselves in zoos in East Asia and North America and in their ancestral homeland. Collectively, they would not form a geographic population in any meaningful sense and they need not form a coherent breeding population.

It might be argued that the lineage population concept, as defined here, is too generic to be of practical use and that something in addition to ancestry and descent, such as the presence of notable physical trait differences, should be required. Since the purpose of this concept is analytic, requiring such differences would be undesirable; doing so would require the definition of a more generic primary concept. Practically speaking, it is hard to believe that the lineage population concept is too generic for biologists and others to use, when these same researchers commonly employ the even more non-specific *population* terminology.

In line with geographic and breeding concepts, both specific and infraspecific groups, however such groups are conventionally distinguished, can be lineage populations. Thus lineage population describes specific, subspecific, and infrasubspecific – local and widespread, domestic and wild – communities of descent. The reason for this inclusive usage is both pragmatic and epistemic. For one, restricting the concept to infraspecific groups would limit its applicability; doing so would also render it asymmetrical with the geographic and breeding population concepts, which typically apply to both infraspecific and specific divisions. Additionally, there are a number of species concepts which differently delimit species (Wilkins 2010). The result is that the species of one concept can

---

[36] Some subspecies concepts, for example the popular evolutionary taxonomists', allow subspecific taxa to be cut from the ends of a continuum (see, for example: Mayr and Ashlock 1991, 50).

be the infraspecific "populations" of another. Thus, for example, Mayr's biological species concept's *subspecies* and *geographic races* can be Wheeler and Platnick's diagnostic phylogenetic species concept's *species* (Mayr 2000; Platnick 2000). Specifying lineage populations as infraspecific groups would require one to specify a species concept. And doing so would unnecessarily embroil the lineage population concept in the seemingly unending debate on the appropriate definition of species.

2.1. Alternative terms for lineage populations

Sometimes, in context to humans, lineage populations are called *ethnic groups*. The term *ethnic group* ("groupe ethnique") was adopted by Joseph Deniker to refer to communities of individuals delineated in terms of language, religion, and culture and to distinguish these social divisions from zoological ones. According to Deniker, the former communities can be composed of a single zoological group or multiple ones (whether species, varieties, or races), whereas the latter are delineated in terms of somatic resemblance (Deniker 1906, 280-284). However, to avoid the use of the term *race*, *ethnicity* has been adopted to refer to somatically and ancestrally delineated groups (Lieberman et al. 2004; Molnar 2015). This practice seems to have been taken up to some extent and the situation has led to the very confusion which Deniker (1906) was attempting to circumvent (Štrkalj 2005). Owing to both semantic ambiguity and its inapplicability to non-human organisms, the term *ethnic group* is undesirable as a substitute for *lineage population*.

Recently, the term *cluster* has been adopted by some – or adopted as a euphemism for *race* meant roughly in the same way (Kitcher 2007). The problem with *cluster* is that it is ambiguous, since it has several substantially different meanings. A *cluster* can mean the statistical output from a cluster analysis, or it can refer to groups of similar organisms identified by cluster analysis; such clusters can refer to all sorts of groups, such as ones delineated by sex or age (Pigliucci 2013). More often, the term *genetic clusters* refers to organismic groups delineated in terms of ancestry-indexing assemblages of phenotypic and/or molecular characters. These lineage clusters would correspond with lineage populations, insofar as the former reliably delineate the latter, which would only be the case if a sufficient number of characters, with a high phyletic weight,[37]

[37] Phyletic weight refers to the "phyletic information content of the character" (Mayr and Ashlock 1991, 185). Lineage populations can be conceptualized as latent factors, ones which are indexed to one degree of reliability or another by sets of characters. If the

were used. Yet *cluster*, in this lineage cluster sense, has also been taken to refer to lineage populations between which there are historic evolutionary phyletic discontinuities. This is because discontinuities resultant of evolutionary patterns (e.g., vicariance), sampling, or recent migrations allow unsupervised cluster analysis to pick out groups; thus the cluster concept is taken, at times, to imply groups between which there are such discontinuities, and it is assumed that, when dealing with systematic or taxonomic groups, these phyletic discontinuities should result from historic evolutionary processes and not, for example, from sampling.[38] Thus, Templeton (2006, 216) tells us that "an isolation-by-distance model with no genetic clusters at all will still have the appearance of genetic clusters…." According to Templeton (2006), *actual* clusters are groups between which there are evolved phyletic discontinuities. As a result of these ambiguities and different readings, terms such as *genetic clusters* or *clusters* are problematic when it comes to describing what are here called *lineage populations*.

In relation to the intent of the lineage population concept, there is also a metaphysical concern with cluster concepts. This can be illustrated using Glasgow's (2010) racial twin earth thought experiment. Glasgow (2010) asks us to imagine the creation of a racial twin earth in which the species and lineage populations/races thereof are identical in form to those on Earth. Based on this thought experiment, he reasons that lineage cannot be central to the meaning of *race*. Across Earths, Glasgow's twin "Europeans," despite being separate creations, would indubitably be arranged into the same biogenomic cluster as Earth Europids. However, as defined here, they would not form the same lineage population – divisions which are *indexed* by clusters of characters but *defined* in terms of descent. Indeed, being the product of different creation events, Earth and twin Earth Europids would represent separate *species* in the 18th century polygenist sense. It is not clear what they would represent in modern phyletic-based systematics.

Yudell et al. (2016) insist that terms such as *ancestry* and *populations*, and not *race*, should be used to describe "human groupings in genetic studies." For

---

phenotypic or molecular characters analyzed are unreliable indices of "true ancestry," one can get clusters which are discordant with the underlying lineage populations.

38 Thus, Templeton (1998) notes that "when a biological race is defined as a distinct evolutionary lineage within a species, the question of race can *only be answered in the context of the recent evolutionary history of the species*." [Italics added.] By his idiosyncratic concept of *race* and *cluster,* for these to be real, they need to represent lineages between which there are/were historic evolutionary disjunctions.

them, *ancestry* is a "statement about an individual's relationship to other individuals in their genealogical history." It is precisely because the desire is to describe *collections* of individuals related in terms of descent that *ancestry* and "process-based" terms and concepts such as *geographic ancestry*, *genetic ancestry*, and *biogeographic ancestry* – are insufficient. Simply put, *ancestry* does not specify groups of organisms, let alone groups delineated in terms of overall ancestry (as opposed to, for example, maternal ancestry). In genetic studies, one is frequently interested in ancestry *with respect to biostatistical populations of a specific sort* – specifically, ones delineated in terms of genetic relatedness or nearness of descent.

The term *geographic ancestry* is additionally problematic since lineage populations are not delineated with respect to the geographic origins of ancestors, per se, but rather with respect to shared ancestry. Only when individuals from roughly the same geographic regions descend from the same linebred populations do they belong to the same lineage populations. As an example of discordance, according to OMB Directive 15, an *Asian* by US census standards is a "person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, and the Pacific Islands," while *Whites* are "persons having origins in any of the original peoples of Europe, North Africa, or the Middle East" (Registrar,1997). In some sense, these political divisions are delineated in terms of geographic ancestry, yet they do not form coherent lineage populations as South Asians are more ancestrally related to West Eurasians than to East Asians and Pacific Islanders.

A related but more precise term, used in epidemiology and medical research, is *biogeographic ancestry* (*BGA*), a term for which "bio-" qualifies ancestry. Gannett (2014) has detailed the origin of the BGA concept. The term *BGA* was first introduced in Pfaff, Parra and Shriver (2000) to refer to "the component of ethnicity that is biologically determined and can be estimated using genetic markers that have distinctive allele frequencies for the populations in question …." Frudakis and Shriver (2003) describe *BGA* as "the heritable component of 'race' or heritage" and note that it "is relevant on any scale of resolution" from continental groups down. Shriver and Kittles (2004) flesh out the concept. According to them, BGA is the component of personal genetic history indexed by ancestrally informative autosomal markers. Unlike maternal and paternal lineage, indexed by mitochondrial and Y-chromosomal DNA, BGA reflects an individual's overall ancestry with respect to local and continental "population groups" (also called *clusters*, *ancestral groups*, and a*ncestral populations*). It is said to reflect the effects of evolutionary factors, such as "isolation by distance," and barriers that "have all affected human migration and mating patterns in the past," which have shaped the present worldwide distribution of genetic variation. As Gannett (2014) notes, Shriver sees the distribution of "genetic variation among people" as continuous and sees this as precluding racial classifications where *race* is apparently seen – contrary to the 18th to early 20st century infraspecific sense – as implying phyletic discontinuities. The reference groups for determining BGA have been called *BGA groups* (e.g., Frudakis & Shriver 2003; Frudakis 2010), though the term *BGA group* or variants of this is not commonly used in the literature. Instead, the reference groups are typically ambiguously just called *populations*. An example of the method used can be seen in Keating et al. (2013). Reference samples were taken from HapMap 3. Principal component analysis, applied to allelic data, was then used to group individuals from these reference samples into major "descent groups." It is notable that the reference groups are not obviously populations in the geographic or breeding sense. The European one, for example, is a biostatistical class based on individuals from Utah, US and Tuscany, Italy. The commonality of the members, in this case, owes to having descended from a once relatively endogamous, geographically circumscribed breeding population. Delineated this way, they are lineage populations. Yet, in principle, not all lineage populations need to be like this. That is, the lineage population concept is meant to be inclusive. All intrafertile classes – allopatric or sympatric – where members are arranged by propinquity of descent are lineage populations. The BGA group concept, then, represents a narrow specification of the lineage population one.

A reviewer suggested that the term *ancestral population* might be used synonymously with what is here called *lineage population*. This, however, seems not to be the case – see, for example, usages in Savolainen et al. (2002) and Bertoni et al. (2003). Specifically, *ancestral population* is used to mean 'one of the populations from which this population is derived' and has meaning only in relation to *derived populations*. *Ancestral population* is doubly problematic insofar as one desires a synonym for lineage population. First, it would be semiotically incorrect to refer to two derived lineage populations as being *ancestral populations*. For example, one might coherently speak of the ancestral population linking gorillas and modern humans, but it would be misleading to refer to these two derived species-level lineage populations as being, with respect to one another, separate *ancestral populations*. Second, it is not clear if *ancestral populations* necessarily correspond with *lineage populations*, or whether they can alternatively correspond with highly genetically heterogeneous spatial populations, as they apparently do in the case of the *ancestral populations* of Bertoni et al.'s (2003) US Latin Americans.

A better substitute term might be *ancestry population*. However, the precise meaning of *ancestry populations* is unclear. From 1980 to 2005, the term seems to have been predominately used by sociologists to refer to populations delineated with respect to geographic origin. Thus, authors speak of, for example, a "Hispanic" ancestry population, where this refers to a class of individuals 'having Latin American origin.' As noted previously, such region of origin delineated groups could correspond with lineage populations but they obviously need not. Between 2006 and 2015, the term has increasingly been used in population genetic and genetic epidemiological studies. However, one still finds references to groups which are apparently delineated in terms of geographic origin – for example, the "Hispanic ancestry population" of Hulgan et al. (2015). It is possible that the term *ancestry population* has come to mean, for most researchers, lineage population, that is, one of a set of divisions into which members are arranged by propinquity of descent. However, authors are frustratingly unclear. Given this situation, instead of imposing a definition on a presently ambiguous term, it is desirable to coin a new term for analytic purposes and then inquire: "Are *ancestry populations*, as meant in context to population genetic research, lineage populations?"

A term proposed by Fuerst (2015) is *natural divisions*, where this refers to divisions into which organisms are arranged by lineage or propinquity of descent (Kant 1777; Darwin 1859). Unfortunately, this term could be taken as implying natural partitions based on phyletic discontinuities. While there are historic precedents for equating ancestry-based arrangements with *natural classifications*, for some, phyletic discontinuities are a prerequisite for the "naturalness" of biological divisions. Gray (1879, 322), for example, notes that "if the species blended as do the colors of the rainbow… there could be no natural foundation for their classification. The multitude of species would render it necessary to classify them, but the classifications would be wholly artificial and arbitrary." For Gray (1879), a natural biological classification entails natural partitions, where nature, as Darwin puts it, enables us "to divide the existing species into definable groups." Following Gray (1879), one could argue that genealogical arrangements into divisions are "not really" natural in absence of sufficient phyletic disjunctions. Thus, the term *natural division* potentially lends itself to confusion.

This notion of lineage population is not new. It corresponds with Darwin's (1859) species and hereditary varieties as "communities of descent." While it is outside the scope of this paper to show this through textual analysis, the idea recapitulates one historic natural scientific understanding of race.[39] Granting that, at a given time, infrasubspecific genetic populations in the retrospective sense, subspecies, and species are most often separate lineage populations, the concept is convenient in that it allows one to circumvent ongoing debates concerning how to distinguish these entities from one another. Moreover, the concept allows one to discuss domestic genetic breeds[40] under the same term much as the polyseme called *race* once did. The lineage population concept further allows one to tie the previously mentioned concepts to ones, such as BGA group or descent group, used in epidemiology and population genetics.

3. Lineage population concept versus the population lineage concept

As defined here, lineage populations are divisions of organisms into which members are arranged by propinquity of descent. This idea is readily contrasted with geographic and breeding populations, understood as groups where members are arranged in terms of, respectively, spatial proximity and the probability of descendant sharing. This concept primarily views organismic groups horizontally across space, as defined by Stamos (2002), and in line with Darwin's perspective (Stamos 2007). A closely related but subtly different concept is called *population lineage*, which Ereshefsky (1992) defines, with respect to species, as "a single descendant-ancestor sequence of organisms or a group of such sequences that share a common origin." The latter concept places emphasis on descendant-ancestor relationship viewed vertically across time.

---

[39] This might be called the "Darwinian concept" of race, since Darwin most clearly formulated the goal of natural or genealogy-based classifications, ones which had the intent of allowing for inductive potency regarding overall organic resemblance. Following Darwin, *races* were often seen as a natural classification in this sense. Thus, for example, in "African Races," Chatelain (1894) notes that: "Many criteria - such as the color, the hair, the form of the skull - have been proposed as bases for the racial classification of mankind. All have been helpful, but none has proved adequate. All are one-sided and artificial, failing to grasp and follow through its ramifications the principle of genealogy which seems to be essential for a natural classification." For a 20th century exposition of this concept, see Brues (1990, 1-7); compare with that in Hooton (1946, 1-2, 447-449).

[40] Historically, breeds were understood to be groups of animals sharing a common phenotype owing to common ancestry resultant of artificial reproductive isolation and selection; see, for example, the definition of Blyth (1835). However, breeders now make the distinction between "genetic breeds" and "phenotypic breeds," the latter which share a common phenotype but are not necessarily particularly related.

Kevin de Queiroz (1999; 2005; 2011) has most extensively developed a general "population lineage"[41] concept of species, according to which a species is a “separately evolving meta-population lineage” (de Queiroz 2005). He reasons that species are population-level “biological entities whose members propagate themselves to form lineages” (de Queiroz 1999). Thus his species concept is very similar to what Wilkins (2009; 2010) identifies as the common historic species concept, which Wilkins calls the “generative concept”: “species has always been thought to mean the generation of similar form. That is, a living kind or sort is that which has a generative power to make more instances of itself” (Wilkins 2009, 232). De Queiroz sees this as a unifying concept; to be general enough to be so it is radically inclusive (for a *species* concept). Despite this, it is nonetheless stipulated that the groups are “separately evolving” and “meta-populations”.[42] The concept then seems to exclude hybridizing and converging population lineages in addition to local ones. If de Queiroz’s constraints of being “separately evolving” and a “metapopulation” are lifted, we are left with an even more general concept, de Queiroz's *population-lineage* [43], which corresponds with what, in context to natural history, *race*, a polyseme when not a concept, typically seemed to mean in the most general sense, that is, as Gray (1879, 320-321) suggested, a “common designation of any group or collection of individuals whose characters are continued through successive generations, whether permanent variety, subspecies, species, or group consisting of very similar species...”

What is relevant to this discussion is that population lineages need not correspond with lineage populations, whether one is dealing with species or infraspecific divisions. This is because the initial members of a population lineage could be as related to the members of their original parental population lineage as to their descendants. This happens, for example, in the case of chromosomal speciation. For example, Milhomem et al. (2008) find evidence of a cryptic

[41] By “population” he means only a group of organisms.

[42] A term, he tells us, used “to distinguish species, which are traditionally considered to reside at the higher end of the population-level continuum, from populations at the lower end of the continuum, such as demes and family groups” (de Queiroz, 2007).

[43] De Queiroz (1999) indicates that the term *population lineage* can apply to the infraspecific or deme-level: "Lineages at lower levels in this continuum (e.g., demes or deme lineages) often separate and reunite over relatively brief time intervals. Toward the other end of the continuum, lineage separation is more enduring and can even be permanent.... In any case, most authors equate species with lineages toward the latter end of the continuum."

species, the 2n=42 form of *G. carapo* sensu stricto, which is reproductively isolated – the authors speculate due to chromosomal speciation – from sister geographic populations but for which the authors could find no "evident differences in external morphology, meristics and pigmentation between the two forms." Presumably, the first few generations of this cryptic species and the sibling species would not represent separate lineage populations – distinguishable based on overall relatedness – but would, nonetheless, represent separate population lineages.[44] A similar disjunction can be found in historic discussions of *race* understood as constant varieties. Defining *race*, Jean Louis Armand de Quatrefages (1861, 81-82) writes:

> But, add the authors that we are here confronting, the differences between these varieties are not very substantial; they only concern unimportant features, such as size, color, etc. – Even though it would be so, even though these differences would be as insignificant as we would like to say, who cares? From the moment they became constant and they pass on by way of inheritance, they are no less real races.

This concept of (infraspecific) race is clearly in line with the generative concept, or at least an infraspecific version of it. However, as with the 2n=42 form of *G. carapo* senso stricto, members of de Quatrefages's (1861) races need not initially be lineage populations. De Quatrefages's (1861) concept was not an uncommon pre-Darwinian one of infraspecific race. Duchesne (1766, 18), for example, points out that race, "the term employed with reason by Buffon in the *Natural History of Animals*, and which asks to be introduced into that of Vegetable," was needed to describe varieties which propagated their form across generations. Otherwise, "by following Ray's axiom, which cannot consider constant races as varieties, they are named species" which represents an inappropriate classification (Duchesne 1766, 26). In context to strawberries, Duchesne (1766) notes:

[44] A similar point is made by Hausdorf (2011) in his discussion of the genotypic-cluster concept of species: "The prediction that species sooner or later form genotypic and phenotypic clusters can be derived from most species concepts. Thus, this is doubtlessly a useful criterion for delimiting provisional species. However, incipient species might not yet be recognizable as distinct clusters based on a random sample of genetic markers. In the case of peripatric speciation, the peripheral species will initially often form a cluster with neighboring populations of the more widespread species so that the more widespread species does not form a genotypic cluster distinct from the peripheral species."

It is certain today that, if all species are stable, there are also races whose distinctions are constant, although belonging to the same species. The Versailles strawberry that I saw born, and which became the head of a race, puts that fact beyond doubt. Cultivation and other accidental causes do not produce new species, but changes in certain individuals do occur that are perpetuated in their posterity, constituting new races. (Appendix, "Remarques Particulières," pp. 11-21).

We see that *race* as constant variety is explicitly the infraspecific equivalent of Wilkins' generative species concept, discussed above, and like the generative species concept, it is derived from – or more correctly in this case, in context to – Ray's biological species concept, the first modern biological one. The disjunction between the lineage population and population lineage concepts becomes more apparent when examining Duchesne's phylogenetic network of strawberry races. Discussing the birth of "La Race nouvelle," he notes that he observed how the fertilization of a Frutiller female by a male Capiton produced a mix who "will form perhaps a new race" (pp. 223-227). Either the initial head of the Versailles strawberry race, produced by the marriage of le Capiton and le Frutiller, would be of the Versailles strawberry race or its offspring would, assuming that the race continued through either self-fertilization or the incestuous mating of le Capiton and le Frutiller's offspring. In either situation, the initial members of one race would be as related to the members of their original parental race as to their descendent race. Thus, strictly understood, members of constant variety – or generative – races need not form lineage populations.

The reason for the focus on lineage populations is that they are more inductively interesting. To paraphrase Darwin, the physiological and morphological resemblance of lineage populations and their pedigrees generally correspond; thus knowing that two groups are separate lineage populations gives one more inductive leverage than knowing that they are only different population lineages. However, insofar as one is interested in the most inclusive concept of either species or what was called *race,* lineage population would not be it. The lineage population concept is rather, from the perspective of systematics, a specification of a more general concept – what one gets when population lineages are sufficiently linebred.[45]

[45] It would require some interpretive footwork to make the case that *races* typically picked out population lineages. To give an example of the hermeneutic problem, as Douglas (2005) notes, Buffon (1777, 462-463, 479-484) allows his Lapp race – using the word *race* "in the broadest sense" (pp. 462) -- which he also refers to as a "species" (pp. 480-481) to be polyphyletic, with the Lapp sub-races grouped together on account of their supposed

3.1. The lineage population concept versus the race as breeding population

There is substantial semiotic overlap between what is here called *lineage populations* and what were in the 18th and 19th centuries called *races*. This stems from both terms referring to lineage-based groups, whether on the specific or infraspecific level.[46] For example, Georges Cuvier (1831, 73) defines species as "individuals who descend from one another, or from common parents, and those which resemble them as strongly as they resemble one another" and notes that "we only call the *varieties* of a species, those races, more or less different, which may have proceeded from them by generation." His races as species were *lineage populations* and his races as varieties were *population lineages* which could also have been lineage populations. As suggested above, though, there are also substantive inconsistencies. This topic is complicated by the existence of a few different primary historical concepts which disagree in important respects and because, during this time, biological variation was approached from very different systematic perspectives.

---

epigenetically induced overall resemblance. Understood as polyphyletic groups, these would not be population-lineages in Ereshefsky's (1992) sense. Yet, given contemporaneous theory, classifications based on overall resemblance, which Buffon was aiming for, would end up delineating lineage-populations, groups which would be types of population lineages.

[46] Doron (2011, 763) notes that "the concept of 'race' ... refers to genealogically constant differences, inscribed in lines, which make it possible to base a classification. It would be difficult to find an author at the end of the eighteenth, nineteenth century and the first half of the twentieth century who did not accept this definition of the concept of 'race'." This statement needs some amending. *Race* qua *constant variety* and qua *species* was doubly used to refer to genealogically constant differences and to the lines of descent along which these were transmitted. Yet, the term was also used to refer to lines of descent which had no determinate relation to differences, thus, for example, the polygenists' infraspecific *races* as *Linnaean varieties*. The common idea across early usages was lineage, as one might anticipate based on the term's etymology. In contrast, the term *species* was understood to imply "genealogically constant differences" – or different forms which reproduced themselves. Discussing the terms, Smith (2015, 143) cogently notes: "'Species' occupied a distinct but nonetheless partially overlapping semantic field with 'race,' to the extent that the former term, as its etymology suggests, was concerned with the external aspect or appearance of a creature… The principal semantic difference between 'species' and 'race,' where these in fact differed, had to do with the fact that the former focused on physical traits of creatures, while the latter also recalled to mind the lineage or generative series from which – to return to the deepest etymology of the term – these traits flow."

Certain 20th century concepts of race at times diverge more substantially from the lineage population one – specifically, concepts which identify races with breeding populations and do not imply (e.g., by also identifying them with reproductive isolates) lineage-based groups. For example, Strkalj (2000) tells us that race either means breeding population or subspecies in Mayr's sense.[47] He suggests further that the term *race*, as typically used in context to biological anthropology, is redundant with the term *breeding population*. What he means by *breeding population* is not clear, but he references the work of Stephen Molnar (1998).

Molnar (2015, 18), in the sixth edition of the book referenced above, recognizes that, historically, *race* was used to refer to "breeds of domestic animals – their groups membership or descent from a common ancestor." Strangely, despite recognizing the historic relation between the notions of race and descent, lineage does not take a prominent role in the conception. After reviewing a few concepts, he notes that there are numerous definitions which seem to agree little except with regards to two common components: an assumption about geography in the formation of races and a placing of importance on "breeding populations." Conveniently, in his glossary, he defines *breeding population*: a "group of individuals who are potentially interbreeding, who occupy a local area, and who make up a basic unit in our species" (p. 420). In light of his conceptual analysis, he decides to use *race* sparingly to mean "a label we append to a population grouping with some physical characteristics that have some genetic component" (p. 398). By this definition, *races* seemingly could be different breeding and geographic populations which happen to, on average, genetically differ in a number of traits. Read literally, ancestrally-heterogeneous North American and Western European residents could collectively represent two different races. Molnar makes it clear that the intent of this understanding is to "appreciate the fluidity of boundaries due to human behavior" (pp. 317-318) His conception seems to accomplish that task by departing greatly from the typical 18th and 19th century lineage-based understanding of *race*.

Of course, the relation between his sense of *race* and ancestry is noticed. For example, he tells us that to replace the term *races*, "ethnic groups" is "more

---

[47] In discussing subspecies, he refers to Mayr's definition. But then he states that the question is whether one can "make a meaningful grouping within the species *Homo sapiens* above the level of breeding populations." A logical asymmetry is involved in the question because the *breeding population* systematic classification is not a level in the way that Mayr's subspecies category is a taxonomic rank.

and more frequently used today as a substitute term for people presumed to be of different ancestral descent…" (p. 309). Likewise, the relation between his sense of *race* and genetic clusters is recognized, since we are told, after he cites the work of Cavalli-Sforza, that what had "once appeared to be a few simple subdivisions of our species turns out to be numerous small and large clusters of genetic heterogeneous groups" (p. 310). One would think that the recognition of the relation between *race* and ancestry and *race* and groups arranged by phyletic-informative clusters of characters would bring to Molnar's mind a different historic race concept – and lead to a different definition. This appears not to be the case. The objective here is not to pick apart Molnar's discussion but to illustrate, again, how the lineage population concept differs from some contemporary and historical race concepts and how some contemporaneous race concepts are often rather opaque, unspecific, and not clearly related to historic lineage-based ones.

4. Conclusion

Yudell et al. (2016) have recently argued that *race* should be taken out of population genetics. Briefly, they see *race* as problematic because as a polyseme it has inconsistent definitions, and as a network of concepts it refers to a taxonomic categorization "based on common hereditary traits" for the purposes of clarifying "the relationship between our ancestry and our genes" and is "pattern-based." They proposed that alternative terms such as *ancestry* and *population* should be used to describe human groups because "the scientific language of race has a considerable influence on how the public (which includes scientists) understands human diversity."

Their second criticism makes little sense since, as Kant (1788) noted over 200 years ago, "the word is not to be found in a[ny] systematic description of nature," that is, the term *race* does not refer to a formal zoological or botanical taxonomic category. Instead, terms such as *genus*, *species*, and *subspecies/variety* did and do. Historically, in context to natural history, *race* simply referred to lineage. Insofar as taxa, and other groups such as domestic breeds and infrasubspecific varieties, were understood to be these, they were called *races*. Thus one could speak of the human race (lineage), the Tartar race (lineage), or a particular race (lineage) of poultry or strawberries. As for their third criticism, "pattern based" seems simply to mean "class based" – to refer to groups defined in terms of patterns of traits. This is a strange criticism since the authors apparently have no problem with the same types of groups traveling under the name of *genetic population* and presumably *BGA group*. The opposition, in this regards, seems to be not to "pattern-based" concepts, per se, but to a word which connects dense conceptual networks to the sanitized “pattern-based” concepts which they feel more comfortable using.

Regardless, it cannot be denied that *race*, the term, especially presently, suffers from having multiple, inconsistent, and disputed definitions. And yet there is nonetheless something in nature to be described, which the term, however ambiguously, used to. A clearly articulated, unifying concept is needed to fully understand biological variation in line with the Kantian/Darwinian principle for natural history – a concept of specific and infraspecific divisions of organisms arraigned in terms of propinquity of descent. By assigning the neutral term, *lineage population*, to this concept, future misconceptions are avoided. With this analytic concept on hand, it can be investigated if concepts referenced by other more frequently used terms have identical meanings.

Some limitations of the concept are worth reiterating. While the lineage population construct is useful, as testified by the number of studies which employ closely related and overlapping constructs, not all lineage-based variation is captured by the concept. Hausdorf (2011) has made a similar point regarding the genotypic-cluster concept of species in context to discussion of a general species concept. In particular, not all “biological entities whose members propagate themselves to form lineages” (de Queiroz 1999) are captured by the concept. As such, some of the entities which historically were called *races* are unidentified by this concept.

References

Allaby, M. (2013). *A Dictionary of Geology and Earth Sciences*, 4th edition. Oxford Univ. Press.

Aulchenko, Y.S. (2010). Effects of population structure in genome-wide association studies. *Analysis of Complex Disease Association Studies: A Practical Guide* 123.

Beatty, John. (1985). Speak of Species: Darwin's Strategy. In: David Kohn (ed.), *The Darwinian Heritage*, pp. 265-282. Princeton: Princeton University Press.

Beeton, S. (1871). *Beeton's Dictionary of Natural History*. London: Ward, Lock & Tyler.

Bertoni, B., Budowle, B., Sans, M., Barton, S.A. & Chakraborty, R. (2003). Admixture in Hispanics: Distribution of ancestral population contributions in the Continental United States." *Human Biology* 75: 1-11.

Blumenbach, J.F. (1825). *A Manual of the Elements of Natural History*. London: W. Simpkin & R. Marshall.

Blyth, E. (1835). An attempt to classify the “varieties” of animals, with observations on the marked seasonal and other changes which naturally take place in various British species, and which do not constitute varieties." *Magazine of Natural History* 8: 40-53.

Bowler, P.J. (1989). *Evolution: The History of an Idea*. Univ. of California Press.

Bory de Saint-Vincent, J.B. & Marcellin, M.J.B.G (1827). *L'Homme (homo): Essai zoologique sur le genre humain*. Paris: Rey et Gravier.

Boyd, W.C. (1950). *Genetics and the Races of Man: An Introduction to Modern Physical Anthropology*. Boston: Little, Brown & Co.

Brace, C.L. (1964). On the race concept. *Current Anthropology* 5: 313-320.

Broca, P. & Blake, C.C. (1864). *On the Phenomena of Hybridity in the Genus Homo*. London: Longmans, Green & Co.

Brues, A.M. (1990). *People and Races*. MacMillan Publishing Co.

Buffon, G.L.L. (1777). *Histoire naturelle générale et particulière, servant de suite à l'histoire naturelle de l'homme: supplement*, Tome quatrième. Paris: Impr. royale.

Buffon, G.L.L. (1778). *Histoire naturelle générale et particulière, servant de suite à l'histoire naturelle de l'homme: supplément. Tome cinquième*. Paris: Impr. royale.

Chatelain, H. (1894). African races. *Journal of American Folklore* 7(27).

Chater, A.O., Brummitt, R.K. & Ehrhart, F. (1966). Subspecies in the works of Friedrich Ehrhart. *Taxon* 15: 95-106.

Cotgrave, R. (1611). *A Dictionarie of the French and English Tongues*. London: Adam Islip.

Cuvier, G. (1831). *A discourse on the revolutions of the surface of the globe, and the changes thereby produced in the animal kingdom*. Philadelphia: Carey & Lea.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life*. London: Murray.

Darwin, C. (1862). *On the Various Contrivances by Which British and Foreign Orchids are Fertilised by Insects: And on the Good Effects of Intercrossing*. London: J. Murray.

Darwin, C. (1871). *The Descent of Man*, 2 Vols. London: J. Murray.

Darwin, C. (1903). *More Letters of Charles Darwin: A Record of His Work in a Series of Hitherto Unpublished Letters*, Vol. 2. New York: Appleton.

Deniker, J. (1906). *The Races of Man: An Outline of Anthropology and Ethnography*. London: W. Scott.

De Quatrefages, A. (1861). *Unité de l'espèce humaine*. L. Hachette.

De Queiroz, K. (1999). The general lineage concept of species and the defining properties of the species category. In: R.A. Wilson (ed.), *Species: New Interdisciplinary Essays*, pp. 49-89. Cambridge, MA: MIT Press.

De Queiroz, K. (2005). Different species problems and their resolution. *BioEssays* 27: 1263-1269.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology* 56: 879-886.

De Queiroz, K. (2011). Branches in the lines of descent: Charles Darwin and the evolution of the species concept. *Biological Journal of the Linnaean Society* 103(1): 19-35.

Doron, C.O. (2011). *Races et dégénérescence. L'émergence des savoirs sur l'homme anormal*. PhD diss., Université Paris-Diderot-Paris VII.

Doron, C.O. (2012). Race and genealogy: Buffon and the formation of the concept of 'race.' In: E. Casetta & V. Tripodi (eds.), *Making Sense of Gender, Sex, Race, and the Family. Humana Mente Journal of Philosophical Studies* (22): 75-109.

Douglas, B. (2005). Notes on 'race' and the biologisation of human difference. *Journal of Pacific History* 40: 331-338.

Duchesne, A.N. (1766). *Histoire naturelle des fraisiers*. Paris: Didot le jeune.

Ereshefsky, M. (1992). Eliminative pluralism. *Philosophy of Science* 59: 671-690.

Forster, G. (1786). *Something More About the Human Races*, translated by Jon Mikkelsen. SUNY Press, 2013.

Frudakis, T. (2010). *Molecular Photofitting: Predicting Ancestry and Phenotype Using DNA*. Elsevier.

Frudakis, T. & Shriver, M. (2003). Compositions and methods for inferring ancestry. U.S. Patent Application 10/644,594, filed August 19.

Fuerst, J. (2015). The nature of race: The genealogy of the concept and the biological construct's contemporaneous utility. *Open Behavioral Genetics*.

Fujimura, J.H., Rajagopalan, R., Ossorio, P.N. & Doksum, K. (2010). Race and ancestry: Operationalizing populations in human genetic variation studies. In: I. Whitmarsh & D.S. Jones (eds.), *What's the Use of Race? Modern Governance and the Biology of Difference*, pp. 170-183. Cambridge, MA: MIT Press.

Gannett, L. (2014). Biogeographical ancestry and race. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47: 173-184.

Girtanner, C. (1796). *Concerning the Kantian Principle for Natural History*, translated by Jon Mikkelsen. Albany: SUNY Press, 2013.

Glasgow, J. (2010). *A Theory of Race*. Routledge.

Gould, G. (1894). *An Illustrated Dictionary of Medicine, Biology and Allied Sciences*. Philadelphia: P. Blakiston Son & Co.

Gray, A. (1879). *Structural Botany: Or Organography on the Basis of Morphology. To Which is Added the Principles of Taxonomy and Phytography, and a Glossary of Botanical Terms*, Vol. 1. New York: American Book Co.

Hausdorf, B. (2011). Progress toward a general species concept. *Evolution* 65: 923-931.

Hochman, A. (2013). Racial discrimination: How not to do it. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44(3): 278-286.

Hooton, E.A. (1946). *Up from the Ape*. New York: Macmillan.

Hulgan, T., Samuels, D.C., Bush, W., Ellis, R.J., Letendre S.L., Heaton, R.K., Franklin, D.R. et al. (2015). Mitochondrial DNA haplogroups and neurocognitive impairment during HIV infection. *Clinical Infectious Diseases* 61: 1476-1484.

Huxley, J.S. (1939). Clines: An auxiliary method in taxonomy. *Bijdr. Dierk* 27(5): 491-520.

Jevons, W.S. (1874). *The Principles of Science: A Treatise on Logic and Scientific Method*, Vol. 1. London: Macmillan.

Jiménez-Tejada, M.-P., Hódar, J.A., González-García, F. & Naranjo, J.A. (2008). Population and species concepts in conservation biology and their treatment in textbooks of ecology and genetics. In: *Biological Sciences Ethics and Education: The Challenges of Sustainable Development. Proceedings of the BioEd 2008 Conference*.

Kant, I. (1777). *On the Different Human Races*, translated by Jon Mikkelsen. Albany: SUNY Press, 2013.

Kant, I. (1785). *Determination of the Concept of a Human Race*, translated by Jon Mikkelsen. Albany: SUNY Press, 2013.

Kant, I. (1788). *On the Use of Teleological Principles in Philosophy*, translated by Jon Mikkelsen. Albany: SUNY Press, 2013.

Keating, B., Bansal, A.T., Walsh, S., Millman, J., Newman, J., Kidd, K., Budowle, B. et al. (2013). First all-in-one diagnostic tool for DNA intelligence: Genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. *International Journal of Legal Medicine* 127: 559-572.

King, R.C., Mulligan, P. & Stansfield, W. (2013). *A Dictionary of Genetics*, 8th edition. London: Oxford Univ. Press.

Kitcher, P. (2007). Does 'race' have a future? *Philosophy & Public Affairs* 35(4): 293-317.

Lenoir, T. (1980). Kant, Blumenbach, and vital materialism in German biology. *Isis* 71(256): 77-108.

Lieberman, L. (1968). The debate over race: A study in the sociology of knowledge. *Phylon* 29: 127-141.

Lieberman, L. & Reynolds, L. (1978). The debate over race revisited: An empirical investigation. *Phylon* 39(4): 333-343

Lieberman, L., Kirk, R.C. & Corcoran, M. (2000).The decline of race in American physical anthropology. *Anthropological Review* 66: 3-21.

Lieberman, L., Kaszycka, K.A., Martinez, A.J., Yablonsky, F.L., Kirk, R.C., Štrkalj, G., Wang, Q. & Sun, L. (2004). The race concept in six regions: Variation without consensus. *Collegium Anthropologicum* 28(2): 907-921.

Littlefield, A., Lieberman, L., Reynolds, L.T., Azevêdo, E.S., Beals, K.L., Brace, C.L., Garn, S.M. et al. (1982). Redefining race: The potential demise of a concept in physical anthropology [and comments and reply]. *Current Anthropology* 23: 641-655.

Martin, E. & Hine, R. (2008). *A Dictionary of Biology*, 6th ed. Oxford: Oxford University Press.

Maupertuis, P.L.M. de (1745). *Venus physique*. In *Oeuvres*. Lyon: Bruyset, 1768.

Mayden, R.L. (1997). A hierarchy of species concepts: The denouement in the saga of the species problem. In: M.F. Claridge, H.A. Dawah & M.R. Wilson (eds.), *Species: The Units of Biodiversity*, pp. 381-423. London: Chapman & Hall.

Mayr, E. (2000). A defense of the biological species concept. In: Q.D. Wheeler & R. Meier (eds.), *Species Concepts and Phylogenetic Theory: A Debate*, pp. 161-166. New York: Columbia Univ. Press.

Mayr, E. (2002). The biology of race and the concept of equality. *Daedalus* 131(1): 89-94.

Mayr, E. & Ashlock, P.D. (1991). *Principles of Systematic Zoology*. New York: McGraw-Hill.

Mayr, E. & Bock, W.J. (2002). Classifications and other ordering systems. *Journal of Zoological Systematics and Evolutionary Research* 40(4): 169-194.

Milhomem, S.S.R., Pieczarka, J.C., Crampton, W.G.R., Silva, D.S., De Souza, A.C.P., Carvalho, J.R. & Nagamachi, C.Y. (2008). Chromosomal evidence for a putative cryptic species in the Gymnotuscarapo species-complex (Gymnotiformes, Gymnotidae). *BMC Genetics* 9(1): 1.

Molnar, S. (1998). *Human Variation: Races, Types, and Ethnic Groups*. New York: Prentice Hall.

Molnar, S. (2015). *Human Variation: Races, Types, and Ethnic Groups*. New York: Routledge.

Montague, A. (1942). *Race: Man's Most Dangerous Myth.* New York: Columbia Univ. Press.

Mukherjee, D.P. (2007). *Outline of Population Genetics*, Vol. 1. Allied Publishers.

Müller-Wille, S. (2007). Figures of inheritance, 1650–1850. In: S. Müller-Wille & H.J. Rheinberger (eds.), *Heredity Produced: At the Crossroads of Biology, Politics, and Culture, 1500–1870*, pp. 35-42. Cambridge, MA: MIT Press.

Noël, F.J.M. & Chapsal, C.P. (1832). *Nouveau dictionnaire de la langue française*. Paris: Roret.

O'Brien, S.J. & Mayr, E. (1991). Bureaucratic mischief: Recognizing endangered species and subspecies. *Science* 251: 1187-1188.

Pfaff, C.L., Parra, E.J. & Shriver, M.D. (2000). Genetic estimation of biogeographical ancestry. *American Journal of Human Genetics* 67: 221-221.

Pigliucci, M. (2013). What are we to make of the concept of race? Thoughts of a philosopher–scientist. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44(3): 272-277.

Platnick, N.I. (2000). A defense of the phylogenetic species concept (sensu Wheeler and Platnick). In: Q.D. Wheeler & R. Meier (eds.), *Species Concepts and Phylogenetic Theory: A Debate*, pp. 185-197. New York: Columbia Univ. Press.

Prichard, J.C. (1836). *Researches into the Physical History of Mankind*, Vol. 1. London: Sherwood, Gilbert & Piper.

Registrar, Federal (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Registrar* (62): 58781-58790.

Richelet, P. (1757). *Nouveau dictionnaire françois*. Geneve: G. de Tournes.

Robinson, R. (2013). *Lepidoptera Genetics: International Series of Monographs in Pure and Applied Biology: Zoology*, Vol. 46. Oxford: Elsevier.

Rycaut, P. (1668). *The Present State of the Ottoman Empire.* London: Starkey & Brome.

Savolainen, P., Zhang, Y.-P., Luo, J., Lundeberg, J. & Leitner, T. (2002). Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610-1613.

Schaefer, J.A. (2006). Towards maturation of the population concept. *Oikos* 112(1): 236-240.

Schrode, K.M., Ward, J.L., Vélez, A. & Bee, M.A. (2012). Female preferences for spectral call properties in the western genetic lineage of Cope's gray treefrog (Hylachrysoscelis). *Behavioral Ecology and Sociobiology* 66: 1595-1606.

Shriver, M.D. & Kittles, R.A. (2004). Genetic ancestry and the search for personalized genetic histories. *Nature Reviews Genetics* 5: 611-618.

Simpson, G.G. (1961). *Principles of Animal Taxonomy*. New York: Columbia Univ. Press.

Simpson, G.G. (1969). *Biology and Man*. New York: Harcourt, Brace & World, Inc.

Smith, J.E.H. (2015). *Nature, Human Nature, and Human Difference: Race in Early Modern Philosophy*. Princeton: Princeton University Press.

Stamos, D.N. (2002). Species, languages, and the horizontal/vertical distinction. *Biology and Philosophy* 17(2): 171-198.

Stamos, D.N. (2005). Pre-Darwinian taxonomy and essentialism–A reply to Mary Winsor. *Biology and Philosophy* 20: 79-96.

Stamos, D.N. (2007). *Darwin and the Nature of Species*. Albany: SUNY Press.

Štrkalj, G. (2000). Form and race: Terminological concepts for the study of human variation. *Mankind Quarterly* 41: 109-118.

Štrkalj, G. (2005). Ideological historiography in biological anthropology: Deniker, Montagu and 'ethnic group'. *International Journal of Humanistic Studies* 4: 34-44.

Templeton, A.R. (1998). Human races: A genetic and evolutionary perspective. *American Anthropologist* 100: 632-650.

Templeton, A.R. (2006). *Population Genetics and Microevolutionary Theory*. Hoboken, NJ: John Wiley & Sons.

Verspoor, E., Stradmeyer, L. & Nielsen, J.L. (eds.) (2008). *The Atlantic Salmon: Genetics, Conservation and Management*. Oxford: Blackwell.

Virey, J.J. (1837). *Natural History of the Negro Race*. Charleston: DJ Dowling.

Wagner, J.K., Yu, J.-H., Ifekwunigwe, J.O., Harrell, T.M., Bamshad, M.J. & Royal, C.D. (2016). Anthropologists' views on race, ancestry, and genetics. *American Journal of Physical Anthropology* 162: 318-327.

Wallace, A.R. (1864). The origin of human races and the antiquity of man deduced from the theory of natural selection. *Journal of the Anthropological Society of London* 2: clviii-clxxxvii.

Waples, R.S. & Gaggiotti, O. (2006). Invited Review: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15: 1419-1439.

Weiss, K.M. & Lambert, B.W. (2011). When the time seems ripe: Eugenics, the Annals, and the subtle persistence of typological thinking. *Annals of Human Genetics* 75: 334-343.

Wilkins, J.S. (2009). *Species: A History of the Idea*, Vol. 1. Berkeley: Univ. of California Press.

Wilkins, J.S. (2010). How many species concepts are there? *The Guardian*, London.

Wright, S. (1978). Evolution and the Genetics of Populations, Volume 4. Variability Between and Among Natural Populations. Chicago: University of Chicago Press.

Yudell, M., Roberts, D., DeSalle, R. &Tishkoff, S. (2016). Taking race out of human genetics. *Science* 351: 564-565.

MANKIND QUARTERLY 2017 57:4 632-658

Group Differences in Adaptive Behavior: Evidence from the USA

Harrison Kane*
Bang Saen, Thailand

Nawithara Foythong
Bangkok, Thailand

*Corresponding author. Email: hdkane@hotmail.com

As a psychoeducational construct, adaptive behavior refers to an individual's independent display of behaviors associated with meeting his/her daily personal and social needs, including behaviors expected in domestic and social environments. Since the 1960's, adaptive behavior has been perceived as a viable alternative to standardized intelligence tests, especially in multicultural contexts. The present study compares the adaptive behavior of Whites and Blacks, using three large normative samples drawn from the standardization of the Adaptive Behavior Assessment System-II. The ABAS-II is an ideal subject for the study of race differences, as it provides a psychometrically solid measure of the spectrum of adaptive behaviors. Findings of multivariate analyses indicate that the adaptive behaviors of Blacks and Whites are largely similar among young children, from infancy to 5 years of age. After the age of 5 years, race differences in adaptive behavior emerge and become more pronounced with age. Appreciable race differences are noted for children, adolescents, and adults. Across the span of adaptive behaviors measured by the ABAS-II, most differences favor Whites. Possible mechanisms for these observed differences are offered, as well as suggestions for future investigation.
Keywords: Adaptive Behavior, Group Differences, Child Development.

Before the advent of formal intelligence testing around the turn of the 20th century, what is now termed adaptive behavior was used to denote and describe an individual's capacity and functioning (Tasse et al., 2012). Social competency, adherence to social norms, positively fending for one's daily life, adaptability to the environment, coping with the demands of everyday life, and social adjustment served as indirect measures of intellect. In his *Development of Intelligence of Children*, Binet (1908) captured this sentiment when he wrote that "An individual is normal if he is able to conduct his affairs of life without having need for the supervision of others, if he is able to do work sufficiently, remunerative to support his own personal needs, and finally if his intelligence does not unfit him for the social environment of his parents" (p. 88). Individuals were deemed lacking/handicapped to the extent they were unable to meet these implicit societal expectations.

In the United States, superintendents of public institutions tasked with the care of the mentally handicapped collaborated to form the Association of Medical Officers of American Institutions for the Idiotic and Feeble Minded in 1876. The association's primary purpose was to discuss "all questions related to the causes, conditions, and statistics of idiocy" (Scheerenberger, 1983). The organization's name was changed to the American Association for the Study of Feeble Minded in 1906, to the American Association on Mental Deficiency (AAMD) in 1933, to American Association on Mental Retardation (AAMR) in 1987, and finally to the American Association on Intellectual and Developmental Disabilities (AAIDD) in 2007. As caring for those with intellectual disabilities assumed greater moral imperative, definitions of mental retardation became more nuanced, sensitive, and descriptive. An important step in refining this definition was taken by the AAMD when mental retardation was characterized as "subaverage general intellectual functioning which originates during the developmental period and is associated with impairment in adaptive behavior (Heber, 1961, p. 3). This definition, with three components (subaverage intelligence, impaired adaptive behavior, and a developmental origin) became widely accepted among service providers and professionals. Importantly, Heber conceptualized adaptive behavior and intelligence as related, but distinct constructs. Subsequent AAMD/AAMR/AAIDD definitions are essentially minor refinements of Heber's 1961 definition.

Heber's inclusion of adaptive behavior in the clinical definition of mental retardation created an immediate need to develop tests that objectively measure the construct. At the time of Heber's proposal, the only test available that measured anything resembling adaptive behavior was the Vineland Social Maturity Scale (VSMS; Doll, 1936), a straightforward rating scale consisting of

117 items. Strictly speaking, because of its limited scope, the VSMS was not a true measure of adaptive behavior. Consequently, the inclusion of adaptive behavior as a diagnostic criterion inspired a proliferation of test development activities in this area during the 1960s (Nihira, 1999). Chief among these activities was a joint research project conducted by the National Institute of Mental Health and the AAMD. The sponsored project culminated in the publication of the first standardized test of adaptive behavior, the *Adaptive Behavior Checklist* (Nihira et al., 1968). Since then, a number of additional tests have been published, with each test attempting to capture evolving definitions and conceptions of the adaptive behavior construct.

Generally speaking, adaptive behavior refers to an individual's independent display of behaviors associated with meeting his/her daily personal and social needs, including behaviors expected in domestic and social environments (Oakland & Harrison, 2008). Adaptive behaviors include the realm of behaviors people are expected to display independently each day. These behaviors constitute the building blocks needed for personal independence and socialization, including the need to lead a normal and rewarding life at home, school, work, and the community. Thus, they are described as functional and authentic life activities (Bagnato, Neisworth & Pretti-Frontczak, 2010). Children, youth and adults who display various special needs (e.g., autism & developmental disability) frequently display deficits in one or more functional adaptive behaviors, prompting special educators and rehabilitation providers to target adaptive skills as the focus of restoration (Ditterline & Oakland, 2009). Adaptive behavior is an obvious and direct reflection of an individual's effective functioning and personal independence. Similar to intelligence, the ubiquity of adaptive behavior often means that it is taken for granted as an important explanative source of individual and group differences (Brand, Constales & Kane, 2003).

As a construct with important applications to education and psychology, adaptive behavior gained considerable acceptance in the 1960s. The embrace of adaptive behavior had its provenance in the misguided view that intelligence tests were biased against Blacks. The controversy has its origins in the consistent finding that Blacks score, on average, 15 points (one standard deviation) lower than their White counterparts on traditional intelligence tests (Jencks & Phillips, 1998; Jensen, 1998; Lynn, 2008; Osborne & McGurk, 1982; Rushton, 1997, 2001). Basically, as the argument goes, intelligence tests are necessarily biased because they are embedded in the dominant (i.e., White) culture, and fail to take into account subordinate (i.e., Black) cultural norms (Jensen, 1980). Eschewing the pejorative notion of "cultural deprivation", scholars coined the term "6-hour retarded child" to describe Black children who appear to struggle with the intellectual demands of school, but function normally at home and in the community. The debate intensified considerably when Nobel laureate William Shockley openly defended the position that race differences in intelligence may have a genetic provenance and that an active program of eugenics was needed to forestall an inevitable devolution of society (Pearson, 1992). The controversy was further stoked with the publication of Jensen's 1969 *Harvard Education Review* article, "How Much Can We Boost IQ and Scholastic Achievement?" in which he questioned the empirical effectiveness of remedial education.

The debate of race differences in intelligence continued throughout the 1970's, accompanied by public condemnation from the Association of Black Psychologists. Notably, in 1979 the state of California legally barred public schools from using standardized IQ tests for determining the eligibility of academically struggling black students for special education services. Siding with Black parents and others who sought to stop the practice, the court determined in *Larry P. v. Riles* (1979) that commonly used IQ tests (e.g., WISC) were racially and culturally biased, resulting in a disproportionate number of Blacks being wrongly labeled as mentally retarded. In 1981, Harvard geologist Stephen Jay Gould authored *The Mismeasure of Man*, a castigation of psychometric intelligence and the statistical techniques used to estimate individual and group differences. *The Mismeasure of Man* was met with uncritical public approbation, earning a number of awards from associations in literature and the humanities (e.g., National Book Critics Circle, American Educational Research Association).

Gould's thesis was easily dismissed by established psychologists, including Arthur Jensen (1982), John B. Carroll (1981), Lloyd Humphreys (1983), and Hans J. Eysenck (1998), noting that his treatise was best characterized as propaganda rather than science. The debate of race differences further escalated in the public consciousness with the publication of Herrnstein & Murray's *The Bell Curve* (1994), which documented the importance of IQ as a construct with far-reaching social consequences. In response, in 1995 the APA's Board of Scientific Affairs (Neisser et al., 1996) issued a position paper on intelligence stating that "there was urgent need for an authoritative report on these issues—one that all sides could use as a basis for discussion" (p. 77). Although nonpartisan and accessible to the general public, the APA report did little to stem the mainstream perception that IQ tests are inherently biased against Blacks. Despite pervasive and open access to information, there is the continued belief that IQ tests are biased against non-Whites. The gulf between accepted public sentiment and expert opinion is so vast and abiding that some scholars have suggested it can be attributed only to deliberate misinformation from the media and scholars who value political correctness over scientific correctness (Gottfredson, 2005, 2007; Snyderman & Rothman, 1990).

With the prevailing media, legal, and public opinions antagonistic toward IQ tests, many professionals (e.g., teachers, psychologists, and clinicians) embraced adaptive behavior as a more palatable alternative to intelligence. Adaptive behavior was seen as socially relevant, unbiased, and a more direct predictor of real-world success. As a result, adaptive behavior is now a necessary and integral component of the assessment to legally classify infants, children and adults as developmentally/intellectually disabled (American Psychiatric Association, 2013).

Amazingly, a review of the available literature reveals no studies devoted to a targeted comparison of the collective adaptive behavior of Blacks and Whites. Occasionally, race comparisons in specific aspects of adaptive behavior are mentioned in the psychology literature, usually only as an afterthought to the primary question under investigation. For example, Bailey and Richmond (1979) reported that Blacks demonstrate better social skills than Whites among intellectually disabled children. Of the dozen or so independent studies that report the separate performance of Whites and Blacks on specific aspects of adaptive behavior, the instrumentation is rather poor, often based exclusively on subjective observation (e.g., Collins & Nowicki, 2001; Matson, 2008). Among the studies that use published tests, the instruments often display questionable standardization, reliability, and validity (Leland et al., 1967; Nihira, Leland & Lambert, 1993). Studies also suffer from restricted age ranges in the sample, as well as an emphasis on exceptionality/maladjustment, rather than typical/normative performance of the population (Adams, McIntosh & Weade, 1973). In addition, much of the previous research is intrinsically plagued by a lack of consensus regarding the theoretical nature of adaptive behavior (i.e., a general factor vs. multiple independent factors) (Bruininks, McGrew & Maruyama, 1988). Basically, the lack of consensus means that different tests of adaptive behavior offered different conceptions of the construct. Thus race differences, if found at all, were inconsistent across studies. In fact, a review of past and current technical manuals of the most popular tests (e.g., Vineland Adaptive Behavior Scales) finds that race comparisons are not even reported. One obvious reason why group differences are not disseminated is that if Blacks are known to perform routinely below Whites, the test may be perceived as biased and fall out of favor, as was the case with IQ tests. Consequently, publishers consciously avoid this line of research in order to maximize sales and profits (T.D. Oakland, personal communication, April 4, 2013). Moreover, the recent history of psychology has indicated that any scholar who dares publish research that illuminates race differences is likely to suffer professional ostracism (Rushton & Jensen, 2008; Tucker, 2009). This lack of purposeful investigation often leaves clinicians and educators with the untested notion that Blacks and Whites are equal in their adaptive performance.

The purpose of the present study is to address the conspicuous omission in the research literature by examining Black-White differences across the established spectrum of adaptive behavior by using standardization data drawn from the Adaptive Behavior Assessment System-II (Harrison & Oakland, 2003a). This will be the first study to examine White-Black differences across the entire range of defined adaptive behaviors. By doing so, clinicians and professionals can better understand the nature of adaptive behavior, as well as its implications for the assessment of children, adolescents, and adults of different races. Additionally, any demonstrable group differences in adaptive behavior would lend convergent validity to the notion that observed Black-White disparities are not statistical artifacts, but differences that are real and have material social relevance.

Method

Participants

Archival data from the Adaptive Behavior Assessment System-II (ABAS-II) were used in this study, eliminating the need for informed consent and review board approval. The ABAS-II standardization data were collected in 107 US cities between 1998 and 2002, are representative of the 2000 U.S. census, and meet and/or exceed all extant guidelines proposed by the American Educational Research Association's Joint Committee on Standards for Educational and Psychological Testing (1999), and International Test Commission (2001) for test development.

Data on 5,286 participants, from birth through 90 years of age, were used in this study. The sample was comprised of 2,647 females and 2,639 males, with average ages of 17.11 and 16.89 years, respectively. The sample included 4,270 Whites (80.8%) and 1,016 Blacks (19.2%). One hundred ninety-four individuals received special education services. For the participants aged birth to 18 years, parents and teachers served as raters/respondents. For the adult sample (above 18 years of age), participants were rated by parents, teachers or family members (e.g., spouse) familiar with the everyday typical behavior of the individual. Details of the sample and forms are presented in Table 1.

Table 1. *Description of the three ABAS-II samples.*

Test form	# Items	Whites				Blacks				Total
		Males	%	Females	%	Males	%	Females	%	
Birth - 5 yrs	241	806	42.0	720	37.6	198	10.4	192	10.0	1,916
5 – 18 yrs	216	766	35.8	910	42.5	216	10.2	246	11.5	2,138
18 – 90 yrs	239	568	46.1	500	40.6	85	6.9	79	6.4	1,232
Totals		2,140	40.5	2,130	40.3	499	9.4	517	9.8	5,286

Instrumentation

The ABAS-II allows for a comprehensive estimate of age-appropriate behaviors necessary for people to function independently, safely, and appropriately in everyday life. The ABAS-II provides an assessment of adaptive behavior consistent with the American Association on Mental Retardation's (1992) and the American Association on Intellectual and Developmental Disability's (2002) definitions together with those from the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorder's 4th (1994) and 5th (2013) editions: a general adaptive composite (GAC), three composite domains (Conceptual, Practical, Social), and ten adaptive behavior skills (Communication, Functional Academic, Health and Safety, Home Living, Leisure, Self-Care, Self-Direction, Social, Motor, and Work Skills). Each skill area represents a conceptually distinct aspect and component of adaptive behavior. Communication, Functional Academics, and Self-Direction skill areas comprise the Conceptual domain. Social and Leisure skill scores combine to make the Social domain. Self-Care, Home or School Living, Community Use, Health and Safety, and Work comprise the Practical domain. The Work Skill area is optional for adults and for youth of working age. Also, there is a Motor Skill area that is part of the GAC for young children, but is not part of any of the domain scores. For clinicians, a prime strength of the ABAS-II is that it is the only standardized test to measure all accepted dimensions of adaptive behavior.

The rating forms of the ABAS-II provide for the measurement of adaptive skills of individuals across multiple environmental settings including home, school, daycare, community, and work. Each rating form is completed independently by the respondent or maybe read to the respondent if he/she does not possess the reading skills required to complete the form independently. Raters familiar with the examinee score each item using a 4-point Likert-type scale. The choices are: is not able, never or almost never when needed, sometimes when needed, and always or almost always when needed. Examples of items representing each of the ABAS-II skill areas are offered in Table 2. Standards for the ABAS-II's development, administration, norms and scoring are detailed elsewhere (Harrison & Oakland, 2003a).

Table 2. *Description of the Adaptive Behavior Assessment System.*

Domain	Skill Area	Description	Example items
Conceptual	Communication	Speech, language and listening skills needed for communication with other people, incl. vocabulary, responding to questions, conversation skills, etc.	Names 10 or more familiar objects. Ends conversations appropriately. Uses up-to-date current information to discuss current events.
	Functional Academics	Basic reading, writing, math, and other academic skills needed for daily independent functioning, incl. telling time, measurement, writing notes/email, etc.	Reads his/her own written name. Sings the alphabet song. Makes correct change for a purchase
	Self-Direction	Skills needed for independence, responsibility, self-control, incl. starting and completing tasks, keeping a schedule, following time limits, following directions, etc.	Stops a fun activity, without complaint, when told time is up. Controls temper when disagreeing with friends. Completes large school projects on time.
Social	Leisure	Skills needed for engaging in and planning leisure and recreational activities, incl. playing with others, engaging in recreation at home or in the community, following rules in a game, etc.	Plays alone with games, toys, and other fun activities. Waits his/her turn in games. Asks to read from a favorite book.
	Social	Skills needed to interact socially and get along with people, incl. having friends, showing and recognizing emotions, assisting others, using manners, etc.	Says "Thank you" when given a gift. Laughs in response to jokes or funny stories. Shows sympathy for others when they are upset or sad.

Domain	Skill Area	Description	Example items
Practical	Self-Care	Skills needed for personal care, incl. eating, dressing, bathing, toileting, grooming, etc.	Buttons his/her own clothing. Uses public restrooms alone. Holds and drinks from a sipping cup.
	Home Living	Skills needed for basic care of a home or living setting, incl. cleaning, straightening, helping with household tasks, care of personal possessions, etc.	Turns television off/on. Wipes up spills. Keeps personal belongings clean.
	Health & Safety	Skills needed for protection of health and to respond to illness and injury, incl. following safety rules, using medicines, showing caution, etc.	Tests hot food before eating. Carries scissors safely. Independently bandages a small cut or wound.
	Community Use	Skills needed for functioning in the community, incl. use of community resources, shopping skills, getting around in the community, etc.	Asks to visit a favorite place in the community (e.g. park). Can use a pay phone. Orders his/her own meals when eating out.
	Work	Skills needed for successful functioning and holding a job in a work setting, incl. completing work tasks, working with supervision, and following a work schedule.	Returns to work willingly after taking a break. Cares properly for work equipment/supplies. Shows up on time for work.
	Motor*	Basic fine and gross motor skills needed for locomotion, manipulation of the environment, and the development of more complex activities such as sports, incl. sitting, walking, kicking, etc.	Shakes rattles or other toys. Uses scissors to cut in a straight line. Stands up from a sitting position.

*The Motor skill area is used to calculate GAC, but not any adaptive domain.

The ABAS-II provides age-based norm-referenced standard scores consistent with those used in most standardized tests based on a normal distribution with a mean of 100 and a standard deviation of 15 for the three domains and the GAC. The measure's scaled scores have a mean of 10 and a standard deviation of 3. Thus, the ABAS-II is linked easily to established and preferred tests used in clinical assessment and the diagnosis of children with exceptionalities, including those with an intellectual disability, autism spectrum disorder, and specific learning disabilities (Ditterline et al., 2008; Ditterline & Oakland, 2009; Harrison & Oakland, 2000, 2003b; Oakland & Daley, 2013; Woolf, Woolf & Oakland, 2010).

The Buros Sixteenth Mental Measurement Yearbook (Spies & Plake, 2005) describes the ABAS-II as developed from sound theory and empirical methods, providing norms from a sufficiently representative and large sample to make meaningful inferences of the population. Reliability studies conducted as part of the standardization process provide evidence of a high degree of internal consistency. Most of the skill area internal consistency coefficients are .90 or greater. The average internal consistency coefficient for the standardization sample's General Adaptive Composite (GAC) ranges from .98 to .99. As well as presenting internal consistency data, the ABAS-II manual lists correlations among the 10 skill areas. Consistent with the theoretical structure of the ABAS-II, average correlations among the skill areas are in the moderate range (.40s to .70s), suggesting related, yet independent skills. Data to support the test's reliability and validity are described as being "impressive" (Burns, 2005; Rust & Wallace, 2004). The theoretical structure and interpretive framework of the ABAS-II is resilient across raters, settings, gender, exceptionalities, clients, and cultures (Chen et al., 2013; Harrison & Oakland, 2003a; Lopata et al., 2012; Wei, Oakland & Algina, 2008). Moreover, the ABAS-II has a number of translations that encourage its use in advanced, emergent, and developing countries (e.g., Cambodia, Thailand, Brazil, Mexico, Taiwan, China & Russia). Consequently, the ABAS-II has become one of the preferred instruments to measure adaptive behavior in international settings. Figure 1 provides a graphical representation of the ABAS-II's theoretical and interpretive structure

Analysis

Analyses proceeded in a straightforward manner. In order to discern developmental trends related to race and other group differences, three age groups were identified: preschool (birth to 5 years), children and adolescents (5 to 18 years), and adult (18 to 90 years). For each age group, univariate and multivariate statistics were used to analyze differences between Whites and Blacks for each of the skill areas. Befitting the hierarchical structure of the ABAS-II, the broad adaptive domains and GAC were analyzed as well. Finally, the method of correlated vectors was applied to each age group to determine any association between W-B differences and the general factor of adaptive behavior (*g*ab).

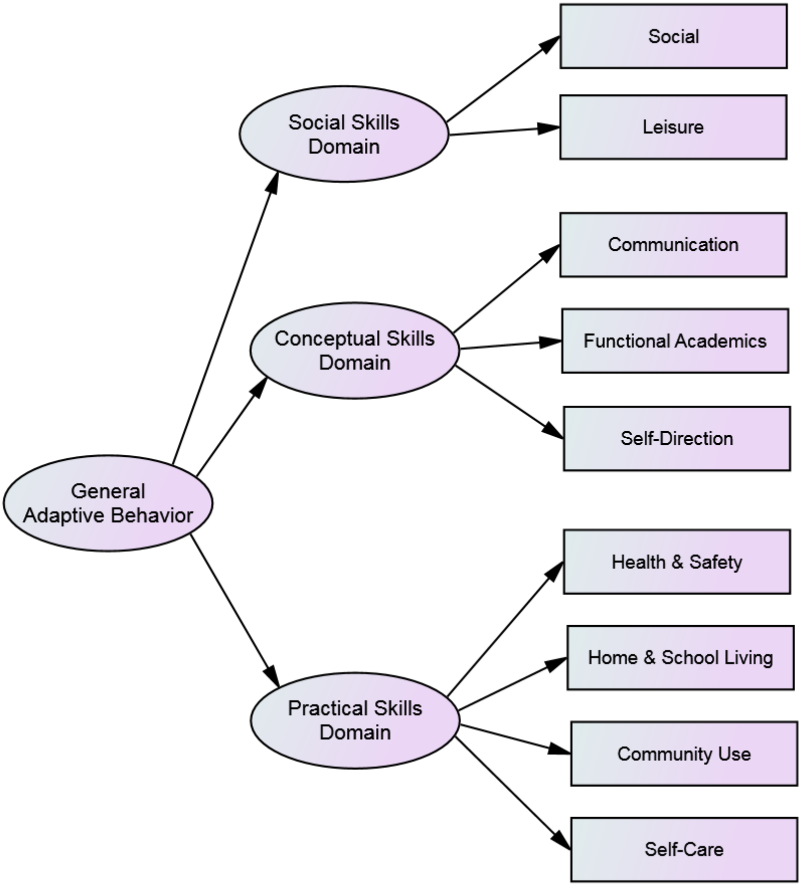

Figure 1. Theoretical Structure of the Adaptive Behavior Assessment System (ABAS)

Results

For the infant/young children standardization sample (birth to 5 years of age), initial multivariate analyses of the ten skill areas yielded statistically significant differences regarding race, Pillais' Trace = .04, F (10, 1513) = 6.45, p<.001. Consequently, a review of the individual subtests seemed warranted (see Table

3). To assist in comparison and interpretation, observed mean differences were standardized in terms of each test's standard deviation for Whites (Hedge's *g*) and then corrected for attenuation and potential bias by dividing each difference by the square root of the particular test's reliability coefficient. Upon review, no discernible pattern of differences emerges. Three comparisons among skills areas prove to be statistically significant (Communication, Community Use, Self-Direction). Of the ten skills areas, four favor Whites (Communication, Leisure, Self-Direction, and Motor), with the remaining skills favoring Blacks (Community Use, Functional Academics, Home Living, Health & Safety, and Self-Care) or offering no apparent difference (Social). Based on the benchmarks suggested by Cohen (1988), the effect sizes of W-B differences are best described as small, ranging from .00 (Social) to .22 (Self-Direction), with an average effect size across adaptive skills of .10. Predictably, effect sizes for the adaptive domains are correspondingly small in magnitude, on the order of .12. Finally, the effect size associated with general adaptive functioning (GAC) is negligible, at .02. In other words, the distributions for Whites and Blacks overlap about 98%. We can conclude that for infants and young children, the adaptive behaviors of Whites and Blacks are very similar, with only minor differences across skills areas, domains, and general adaptive performance.

Table 3. *Comparisons of ABAS-II subtests, domains, and General Adaptive composite across race, for ages birth through 5:0 years, and subtest loadings on the common factor (*gab*). *, statistically significant race difference at p<.05. N = 1,526 Whites, 390 Blacks.*

ABAS-II-II Skill/Domain	White Mean±SD	Black Mean±SD	Effect size	White *gab*	Black *gab*
Communication	9.72±3.29	9.20±3.53*	.17	.81	.77
Community Use	9.39±3.16	9.83±3.76*	-.15	.86	.83
Functional Academics	9.54±3.10	9.61±3.78	-.02	.71	.76
Home Living	9.60±3.10	9.97±3.59	-.12	.82	.78
Health & Safety	9.49±3.17	9.67±3.54	-.06	.88	.86
Leisure	9.81±3.15	9.53±3.52	.09	.85	.84
Self-Care	9.64±3.06	9.93±3.60	-.10	.78	.74
Self-Direction	9.99±3.03	9.36±3.60*	.22	.91	.86
Social	9.69±3.10	9.70±3.50	.00	.89	.84
Motor	9.79±3.02	9.52±3.86	.09	.76	.74
Social	96.99±16.85	96.21±16.91	.05		
Conceptual	97.36±17.09	94.57±17.53*	.17		
Practical	95.86±16.89	98.10±17.50*	-.14		
General Adaptive	96.24±16.44	95.92±17.24	.02		

Turning to the child/adolescent standardization sample (5 to 18 years of age, Table 4), test score differences between the race groups were evident with Whites significantly outperforming Blacks across the entire spectrum of nine adaptive skills, the three adaptive domains, and the general adaptive domain (Pillais' Trace = .05, F (9, 1955) = 12.02, p<.001). The effect sizes of W-B differences are best described as moderate, ranging from .28 (Community Use) to .78 (Self Direction), with an average effect size across adaptive skills of .46. Effect sizes for the adaptive domains are on the order of .40, and the effect size associated with general adaptive functioning is .42. With an effect size of this magnitude, the distributions of Whites and Blacks overlap about 84%. As a result, about 68% of Whites score higher than the average Black individual. This study is the first to examine W-B differences across the entire spectrum of adaptive behavior. Consequently, there are no previous studies that may serve as a basis for comparison. However, these findings suggest that W-B differences materialize during childhood and pervade across all aspects of adaptive behavior depicted by the ABAS-II.

Table 4. *Comparisons of ABAS-II subtests, domains and composite across race, for ages 5:00 through 17:11 years, and subtest loadings on the common factor (*gab*). All comparisons are statistically significant at p<.05. N = 1,676 Whites, 462 Blacks.*

ABAS-II-II Skill/Domain	White Mean±SD	Black Mean±SD	Effect size	White *gab*	Black *gab*
Communication	9.03±3.28	7.52±3.82	.48	.85	.87
Community Use	9.00±3.21	8.16±3.56	.28	.77	.82
Functional Academics	9.37±3.09	8.06±3.30	.44	.73	.92
Home Living	9.39±3.20	8.25±3.65	.37	.86	.89
Health & Safety	9.31±3.04	8.23±3.58	.38	.83	.84
Leisure	9.66±2.94	8.38±3.33	.46	.85	.84
Self-Care	9.63±2.95	8.43±2.99	.44	.67	.64
Self-Direction	9.84±3.37	7.32±3.83	.78	.85	.86
Social	9.57±3.07	8.30±3.59	.43	.84	.87
Social	99.88±15.77	92.74±16.16	.46		
Conceptual	98.63±16.03	90.71±16.19	.51		
Practical	99.16±15.56	93.32±16.82	.38		
General Adaptive	99.08±16.45	92.44±16.72	.42		

Regarding adults 18 to 90 years of age (Table 5), W-B differences are apparent and sizeable (Pillais' Trace = .05, F (9, 1150) = 5.85, p<.001). As with children and adolescents, all differences demonstrate that Whites have a significant advantage over Blacks. The average effect size across skill areas is large, averaging .84. This time, the largest W-B difference is found among the skills and competencies demanded in work environments (Hedge's g = 1.08). The smallest W-B difference pertains to the daily living skills that are typically exhibited at home (Hedge's g = .49). Effect sizes across the three adaptive domains are similarly large, averaging .84, while the magnitude of the W-B difference in overall adaptive behavior is .94. Stated differently, approximately 83% of Blacks perform below the average White individual. Thus, W-B differences among adults are best described as considerable, pervasive, and consistent. This 15 point outcome is quite similar to the oft-reported and infamous finding that Blacks score approximately one standard deviation below Whites on standardized tests of intelligence (e.g., Cattell & Cattell, 1973; Lynn, 2015).

Table 5. *Comparisons of ABAS-II subtests, domains, and composite across race for ages 18:00 through 90:00 years, and subtest loadings on the common factor (gab). All comparisons are significant at p<.05. N = 1,068 Whites, 164 Blacks.*

ABAS-II-II Skill/Domain	White Mean±SD	Black Mean±SD	Effect size	White *gab*	Black *gab*
Communication	10.08±3.36	7.10±4.07	0.93	.89	.92
Community Use	9.82±3.40	6.78±4.21	0.95	.92	.96
Functional Academics	9.93±3.59	6.97±3.76	0.87	.91	.95
Home Living	9.71±3.32	8.10±4.17	0.49	.81	.88
Health & Safety	9.72±3.36	7.36±3.48	0.74	.91	.92
Leisure	9.87±3.32	7.35±3.99	0.80	.88	.93
Self-Care	9.99±2.98	7.70±3.94	0.79	.83	.90
Self-Direction	9.98±3.41	7.04±4.08	0.91	.92	.94
Social	9.87±3.37	7.14±4.12	0.86	.88	.92
Work	9.99±3.19	6.70±3.19	1.08	.77	.87
Social	98.56±16.62	85.59±16.39	0.81		
Conceptual	98.45±16.86	83.26±17.70	0.94		
Practical	97.66±17.12	84.85±16.50	0.78		
General Adaptive	96.73±17.21	81.19±16.47	0.94		

Although adaptive skills and domains are offered for clinical interpretation, the ABAS-II is first and foremost a measure of overall adaptive behavior, as represented by the GAC (Wei, Oakland & Algina, 2008). A number of independent validity studies have affirmed that the ABAS-II skills areas contribute primarily to a unitary general factor of adaptive behavior (du Plessis, 2015). Therefore, the method of correlated vectors was applied to determine whether the magnitude of W-B differences in adaptive skills is related to loadings on this general adaptive factor. The technique of using correlated vectors to investigate group differences was developed and refined by Jensen in his many investigations of Spearman's hypothesis (e.g., 1985, 1987). Details of the method are outlined in Jensen (1998), and examples of the technique in practice can be found in Lynn and Owen (1994) and Rushton (2002).

Taking on the child/adolescent sample, general adaptive behavior was determined within each race as the first un-rotated principal factor. In each analysis, a single factor provided eigenvalues greater than unity, affirming the ABAS-II is unidimensional for both Whites and Blacks. The general factor (*g*ab) was particularly robust, accounting for 69.2% and 72.6% of the variance for Whites and Blacks, respectively. For each group, this factor was corrected for attenuation by dividing the loadings of the various subtests by the square root of the reliability for the test concerned. Congruence coefficients were very high (0.99) for the White and Black samples, suggesting that to a large extent, the same factor is measured in both race groups. Finally, correlation coefficients between the vectors of standardized W-B differences and each group's *g*ab loadings are .22 and .17 for Whites and Blacks, respectively (NS, $p>.25$). These results affirm that race differences are not significantly related to the degree each skill area draws upon general adaptive behavior.

For the adult sample, general adaptive accounted for 73.8% and 80.3% of the variance in adaptive behavior among Whites and Blacks, respectively. Again, coefficients of congruence are quite high, above .90. Correlation coefficients between the vectors of standardized race differences and loadings on the general adaptive factor are not statistically significant at .09 and .22 for Whites and Blacks, respectively. As with children and adolescents, W-B differences among adults are not associated significantly with general adaptive behavior.

Although no consistent W-B differences were found among infants and young children, the method of correlated vectors was repeated. Again, coefficients of congruence were above .90, with the general adaptive factor explaining 67.4% and 68.9% of the variation in adaptive skills among Whites and Blacks, respectively. Correlations between standardized differences and general adaptive loadings are nonsignificant, at .15 and .13 for Whites and Blacks, respectively. Consequently, it can be concluded that W-B differences in the adaptive behavior of infants, children and adults are largely independent of any variation in the general adaptive factor.

Discussion

Adaptive behavior has garnered considerable acceptance among educators and psychologists as an essential component in the assessment of Blacks and other nonwhite populations. At the time of its publication, the ABAS-II was the first and only published standardized test that provides a direct representation of the AAMD/AAMR/AAIDD definition of adaptive behavior. The findings of this study offer strong support for race differences across the range of adaptive behavior skills measured by the ABAS-II. These differences are not consistent across the lifespan, however. Below the age of 5 years, Whites and Blacks demonstrate essentially the same levels of adaptive skill. Of the ten skills areas measured by the ABAS-II within this age range, only three offer statistically significant differences. Communication and Self-Direction skills favor Whites, while Blacks demonstrate better developed Community Use skills. Although statistically significant, these group differences are rather meager in magnitude, with an average effect size of .18.

By the time children enter school, usually around 5 years of age, race differences materialize across the entire spectrum of adaptive behavior. Differences are now consistent, with each comparison favoring Whites. Furthermore, the magnitude of W-B contrasts increases to an average of .45 across adaptive skills. The emergence of these W-B differences in adaptive behavior closely follows the pattern observed with race disparities in intelligence (Rushton & Jensen, 2005). The race differences in adaptive behavior are palpable through adolescence and into adulthood. Importantly, the magnitude of W-B differences continues to grow in adulthood, increasing to an average of .84 across adaptive skills.

Rushton's 2000 treatise *Race, Evolution, and Behavior: A Life History Perspective* summarized a number of studies that confirmed faster maturation and development for Black infants on various developmental milestones including sitting up, walking, and running. For example, the average age at which Black children walk is 11 months, compared to 12 months for Whites (Rushton, 2000). Lynn (1998) provided further support that the precocity of Black children may be an international phenomenon. In a comparison of South African Blacks and Whites on the *Bayley Scales of Infant Development*, Lynn ascertained that Black infants were significantly more advanced in terms of their mental and motor development during the first 15 months of life. Consequently, the pattern of observed W-B differences identified in the present study, particularly the *lack* of differences for young children, may simply reflect the early precocity of Blacks. Once their precocity fades, Blacks begin to fall behind Whites in their adaptive behavior. Relatedly, the pattern of race differences may reflect the nature of the home environment for infants and young children. For example, parents typically assist in the feeding and clothing of young children. Many studies (e.g., McMurtry, 2013) report that while Blacks are significantly more authoritarian than Whites in their parenting style, both groups are similar in their involvement and supervision. Consequently, many real and genuine race differences in adaptive behavior may be obscured by parental influence and/or support to the young Black child. Simply stated, Black parents may provide support to compensate for poor adaptive skills. This additional parental support contributes to the higher scores for Black infants and young children.

In the present study, race differences emerge at approximately 5 years of age, the time when children typically transition to school settings and are expected to display greater independence. Another conspicuous increase is noted when individuals reach adulthood, leaving public school for a wide range of settings that presumably suit their chosen lifestyles and aptitudes. Simply put, as individuals mature, adaptive differences across race become more pronounced as the demands of the environment become progressively difficult. As with intelligence, the complexity of everyday life and the challenge of independence may illuminate race differences in adaptive behavior (Gottfredson, 1997). A final plausible explanation is that observed group differences in adaptive performance may actually be an artifact related to the intervening influence of intelligence.

Several scholars (e.g., Scott & Sinclair, 1997) have noted that race differences in IQ appear quite early in development. For example, comparisons of Black and White 3-year-olds in the Stanford-Binet IV standardization sample affirm the oft-documented one standard deviation difference in mean IQ after being matched on gender, birth order, and maternal education (Peoples, Fagan & Drotar, 1995). Previous research reveals no consensus on the actual correlation between adaptive behavior and IQ. Correlations with IQ may vary widely, because different tests capture contrasting aspects of adaptive behavior. Lambert (1981) found correlations between the WISC-R and the AAMD Adaptive Behavior Scale ranging from .18 to .63. Similarly, Harrison (1987) reviewed the available research and found correlations ranging from .03 to .91, depending on the examinee and test. For the ABAS-II, studies conducted during its development and standardization proffer an average correlation between IQ (as measured by the Wechsler Scale appropriate to the age of the examinee) and the GAC of +.58 (Harrison & Oakland, 2003b). Stated differently, IQ can account for approximately 34% of the observed variation in adaptive behavior. The strength of the association is somewhat expected, inasmuch as successful adaptation requires independent problem solving in a germane social context. For example, riding public transportation requires forethought to schedule the trip, reading comprehension to plan the route, a functional understanding of arithmetic/currency to buy the ticket, and sensitivity to social manners/mores. Certainly, intelligence is implicated in adaptive performance (Harrison, 1987). Some scholars have even suggested that the constructs should merge (Schalock, 1999). The relatively small size of the Black preschool sample precludes a reliable comparison between races at specific ages; therefore, we could not pinpoint the exact age at which race differences in adaptive behavior emerge. Still, W-B differences are palpable and pervasive by five years of age. Therefore, the observed W-B racial disparities in adaptive behavior may be coincidental with the contribution of fundamental group differences in intelligence.

Spearman's hypothesis (1927) states that differences in IQ between Blacks and Whites are due primarily to related differences in general intelligence (Spearman's *g*). Simply stated, tests with higher *g* loadings display larger W-B differences. As a classic example, the Digit Span subtest of the WISC is comprised of two components: Digits Forward and Digits Backward. Digits Forward requires the examinee to repeat a random sequence of one-digit numbers in the order they are presented, starting with two digits and adding another number with each additional item. Digits Backward works exactly the same way, except that digits must be repeated in reverse order. Digits Forward is a straightforward memory task, while Digits Backward requires considerable mental manipulation to arrive at the appropriate response. As such, Digits Backward is much more *g*-loaded than Digits Forward. Studies (e.g., Jensen & Figueroa, 1975) have consistently found that the W-B difference for Digits Backward is nearly twice as large as the difference in Digits Forward. The method of correlated vectors is the most commonly used statistical technique to test Spearman's hypothesis, in which the vector of standardized W-B differences is correlated with the vector of each test's *g* loading. Spearman's hypothesis is consistently confirmed, with correlations usually averaging +.60 (Jensen, 1998). Much like intelligence, when tests of adaptive behavior are submitted to factor analysis, they offer a general factor (i.e., *gab*) that accounts for most of the variance in scores (Bruininks, McGrew & Maruyama, 1988). For the ABAS-II, this study found the general factor is especially robust, accounting for an average of 72% of the variance in test scores across the three age groups. When the vectors of W-B differences were correlated with the skills' loadings on *g*ab, coefficients were universally positive but too small to attain statistical significance, averaging only .17 across the age groups. Therefore, the *g*ab factor makes a meager contribution in accounting for the observed W-B differences in adaptive performance.

Unlike tests of intelligence which require an immediate response under standardized administration, the nature of adaptive behavior allows for the considerable contribution of an examinee's personal style and choice. Consequently, personality and other variables (e.g., motivation) may play a more significant role in accounting for the observed race disparities in performance. In general, compared to Whites, Blacks tend to be more deviant, extroverted, disobedient, disagreeable, deceptive, and aggressive (Herzberger & Dweck, 1978; Levin, 1997; Lynn, 2002; Rushton, 2000; Wilson & Herrnstein, 1985). Moreover, some aspects of personality (e.g., aggression) display heritability estimates that are on par with those associated with Spearman's *g* (Mason & Frick, 1994). For Blacks, the established combination of poor reasoning, impulsivity, and antisocial psychopathy may counterpoise the expression of positive adaptive behavior. Conceivably, as the environment allows for greater personal choice and independence, (without the constant intervening supervision of teachers and parents), Blacks increasingly manifest these underlying maladaptive traits. Obviously, future studies of adaptive behavior should include exogenous variables such as IQ and temperament that may contribute to individual outcomes and group averages.

For decades, clinicians have viewed adaptive behavior as an "unbiased" and "fair" alternative and complement to intelligence tests. Some service agencies in the USA have eschewed traditional intelligence tests entirely in the psychoeducational assessment of Blacks, favoring tests of adaptive behavior and other procedures (e.g., behavioral analysis, Response to Intervention). The embrace of adaptive behavior as an appropriate alternative to intelligence tests is surprising, considering so few published studies have systematically investigated the possibility of racial differences. The findings of this initial study suggest that the uncritical acceptance of adaptive behavior is problematic. Although comparisons find that no consistent racial differences are observed for infants and young children, from the time children enter school, W-B differences are pervasive, palpable, and persistent. In fact, W-B differences increase with age, to an average of nearly one standard deviation in favor of Whites. Regarding the practicalities of assessment, clinicians need to anticipate that racial differences in adaptive behavior are likely to accompany and mirror the divergence associated with IQ. The measurement of adaptive behavior is often associated with the determination of special education eligibility. Consequently, compared to Whites, clinicians should expect significantly more Blacks to be diagnosed with an intellectual impairment/developmental disability. Past research, as well as reality, affirms this supposition. Yeargin-Allsopp et al. (1995) found the prevalence of mild mental retardation was more than twice as high in

Blacks as in Whites (odds ratio = 2.5), controlling for SES, birth weight, maternal age, sex, birth order, and maternal education.

Results of the present study should be interpreted with some reservation because conceptualizations of the adaptive behavior construct are still evolving. Desired adaptive behaviors necessarily change as the expectations and opportunities of society change. Nonetheless, much like revisions of intelligence tests, these changes in measurement are usually superficial updates in content, rather than significant changes in concept. For example, a typical adaptive behavior within the Community Use skills area is "Independently uses a payphone". With the inescapable ubiquity of the internet and mobile networks over the past few decades, payphones are now antiquated and extinct. Consequently, this item is often revised to "Independently uses a cell phone". Even across first-world and developing countries/societies (e.g., USA and Thailand), there appears to be a core of adaptive behaviors that is relevant and predictive of successful and independent functioning in an ever-expanding synchronized global community (Oakland, 2005, 2009; Schalock, 2011). Accordingly, the ABAS has garnered growing acceptance as a clinical instrument in different nations/cultures. Therefore, while the findings of the current study are particular to the USA, similar group differences are likely to be found in many other countries/cultures.

With regard to impoverished and third-world countries, an intriguing possibility is that adaptive behaviors may actually emerge earlier than in more developed/affluent societies, simply because the everyday hardships imposed by poverty and a lack of resources may necessitate greater independence in children. For example, young children in Uganda must fetch water, feed livestock and assist in meal preparation, while children of the same age in the USA have no responsibilities placed upon them. While the official minimum age for marriage in Uganda is 18 years, recent census data indicates that 10% of girls are married and pregnant before the age of 15 years, due to accepted tribal practices (UNICEF, 2016). Clearly, more research is vital to determine if the adaptive behavior construct is applicable across nations that are diverse in race, language, religious values and wealth (e.g., England, Saudi Arabia and Liberia).

Inasmuch as this is the first published study of W-B differences across the comprehensive range of adaptive behaviors, it underscores the need for further research. Replication is needed with different instruments (e.g., Vineland Adaptive Behavior Scales) and groups (e.g., Asians, Hispanics). Future studies should identify and measure the influence of possible covariates (e.g., IQ, SES, personality) as well. In summary, tests of adaptive behavior do not cast light upon the hidden talents of Blacks that are ignored by "biased" IQ tests. Rather, White-

Black differences in adaptive behavior largely follow the trends and patterns that are frequently associated with standardized tests of intelligence.

References

Adams, J., McIntosh, E. & Weade, B. (1973). Ethnic background, intelligence, and adaptive behavior scores in mentally retarded children. *American Journal of Mental Deficiency* 78: 1-6.

American Association on Intellectual and Developmental Disability (2002). *Mental Retardation: Definition, Classification, and Systems of Support*, 10th ed. Washington, DC: Author.

American Association on Mental Retardation (1992). *Mental Retardation: Definition, Classification, and Systems of Support*. Washington, DC: Author.

American Educational Research Association, Joint Committee on Standards for Educational and Psychological Testing, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association & American Psychiatric Association Task Force on DSM-IV (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. Washington, DC: American Psychiatric Association.

American Psychiatric Association DSM-5 Task Force & American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Association.

Bagnato, S.J., Neisworth, J.T. & Pretti-Frontczak, K. (2010). *Linking Authentic Assessment and Early Childhood Intervention: Best Measures for Best Practices*, 2nd edition. Baltimore, MD: Paul H. Brookes Publishing Co.

Bailey, B.S. & Richmond, B.O. (1979). Adaptive behavior of retarded, slow-learner, and average intelligence children. *Journal of School Psychology* 17: 260-263.

Binet, A. & Simon, T. (1908). The development of intelligence in children. *L'annee Psychologique* 14: 1-90.

Brand, C.R., Constales, M. & Kane, H.D. (2003). Why ignore the g factor? – Historical consideration. In: H. Nyborg (ed.), *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, pp 503-530. Boston MA: Pergamon Publishing.

Bruininks, R., McGrew, K. & Maruyama, G. (1988). Structure of adaptive behavior in samples with and without mental retardation. *American Journal on Mental Retardation* 93: 265-272.

Burns, M.K. (2005). Test review of the Adaptive Behavior Assessment System – Second edition. In: B.S. Plake, J.C. Impara & R.A. Spies (eds.), *The Sixteenth Mental Measurement Yearbook*. Lincoln: University of Nebraska Press.

Carroll, J.B. (1981). Reflections on Stephen Jay Gould's *The Mismeasure of Man*. *Intelligence* 21: 121-134.

Cattell, R.B. & Cattell, A.K.S. (1973). *Measuring Intelligence with the Culture Fair Tests*. Champaign, IL: IPAT.

Chen, J.H., Oakland, T.D., Chen, H.-Y. & Iliescu, D. (2013). Cross-national assessment of adaptive behavior in three countries. *Journal of Psychoeducational Assessment* 31: 435-447.

Collins, M. & Nowicki, S. (2001). African American children's ability to identify emotion in facial expressions and tones of voice of European Americans. *Journal of Genetic Psychology* 162: 334-336.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition. Hillsdale, NJ: Lawrence Erlbaum.

Ditterline, J., Banner, D., Oakland, T.D. & Becton, D. (2008). Adaptive behavior profiles of students with disabilities. *Journal of Applied School Psychology* 24: 191-208.

Ditterline, J. & Oakland, T.D. (2009). Relationships between adaptive behavior and impairment. In: *Assessing Impairment: From Theory to Practice*, pp. 31-48. Boston, MA: Springer.

Doll, E.A. (1936). Preliminary standardization of the Vineland Social Maturity Scale. *American Journal of Orthopsychiatry* 6: 283-293.

du Plessis, S. (2015). Adaptive Behaviour Assessment System: Indigenous Australian adaptation model. *European Journal of Social and Behavioral Sciences* 15: 1-15.

Eysenck, H.J. (1998). *Intelligence: A New Look*. New Brunswick, NJ: Transaction Publishers,

Gottfredson, L.S. (1997). Why g matters: The complexity of everyday life. *Intelligence* 24: 79-132.

Gottfredson, L.S. (2005). Suppressing intelligence research: Hurting those we intend to help. In: R.H. Wright & N.A. Cummings (eds.), *Destructive Trends in Mental Health: The Well Intentioned Path to Harm*, pp. 155-186. New York, NY: Routledge.

Gottfredson, L.S. (2007). Applying double standards to 'divisive' ideas: Commentary on Hunt and Carlson. *Perspectives on Psychological Science* 2: 216-220.

Gould, S.J. (1981). *The Mismeasure of Man*. New York: Norton.

Harrison, P.L. (1987). Research with adaptive behavior scales. *Journal of Special Education* 21: 37-68.

Harrison, P.L. & Oakland, T.D. (2000). *Adaptive Behavior Assessment System*. San Antonio, TX: Psychological Corporation.

Harrison, P.L. & Oakland, T.D. (2003a). *Adaptive Behavior Assessment System – Second Edition (ABAS-II)*. San Antonio, TX: Psychological Corporation.

Harrison, P.L. & Oakland, T.D. (2003b). *Technical Report: ABAS-II Adaptive Behavior Assessment System – Second Edition*. San Antonio, TX: Harcourt Assessment.

Heber, R.F. (1961). A manual on terminology and classification in mental retardation. *American Journal of Mental Deficiency*, Monograph Supplement 63.

Herrnstein, R.J. & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life.* New York: Free Press.

Herzberger, S. & Dweck, C. (1978). Attraction and delay of gratification. *Journal of Personality* 46: 215-227.

Humphreys, L. (1983). Review of *The Mismeasure of Man* by Stephen Jay Gould. *American Journal of Psychology* 96: 407-415.

International Test Commission (2001). International guidelines for test use. *International Journal of Testing* 1: 93-114.

Jencks, C. & Phillips, M. (eds.) (1998). *The Black–White Test Score Gap*. Washington, DC: Brookings Institution Press.

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

Jensen, A.R. (1980). *Bias in Mental Testing*. New York: Free Press.

Jensen, A.R. (1982). The debunking of scientific fossils and straw persons. *Contemporary Education Review* 1: 121-135.

Jensen, A.R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences* 8: 193-219.

Jensen, A.R. (1987). Further evidence for Spearman's hypothesis concerning black–white differences on psychometric tests. *Behavioral and Brain Sciences* 10: 512-519.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Greenwood Publishing.

Jensen, A.R. & Figueroa, R.A. (1975). Forward and backward digit span interaction with race and IQ: Predictions from Jensen's theory. *Journal of Educational Psychology* 67: 882-893.

Lambert, N.M. (1981). *AAMD Adaptive Behavior Scale*. Monterey, CA: Publishers Test Service.

Larry P. v. Riles, 793 F.2d 969 (9th Cir. 1979).

Leland, H., Shellhaas, M., Nihira, K. & Foster, R. (1967). Adaptive behavior: A new dimension in the classification of the mentally retarded. *Mental Retardation Abstracts* 4(3): 359-387.

Levin, M. (1997). *Why Race Matters*. Westport, Conn.: Praeger.

Lopata, C., Thomeer, M.L., Volker, M.A., Lee, G.K., Smith, T.H., Smith, R.A., McDonald, C.A., Rodgers, J.D., Lipinski, A.M. & Toomey, J.A. (2012). Feasibility and initial efficacy of a comprehensive school-based intervention for high-functioning autism spectrum disorders. *Psychology in the Schools* 49: 963-974.

Lynn, R. (1998). New data on black infant precocity. *Personality and Individual Differences* 4: 801-804.

Lynn, R. (2002). Racial and ethnic differences in psychopathic personality. *Personality and Individual Differences* 32: 273-316.

Lynn, R. (2008). *The Global Bell Curve: Race, IQ, and Inequality Worldwide*. Augusta, GA: Washington Summit.

Lynn, R. (2015). *Race Differences in Intelligence: An Evolutionary Analysis*. Whitefish, MT: Washington Summit.

Lynn, R. & Owen, K. (1994). Spearman's hypothesis and test score differences between Whites, Indians and Blacks in South Africa. *Journal of General Psychology* 121: 27-36.

Mason, D. & Frick, P.J. (1994). The heritability of antisocial behavior: A meta-analysis of twin and adoption studies. *Journal of Psychopathology and Behavioral Assessment* 16: 301-323.

Matson, J.L. (2008). *Clinical Assessment and Intervention for Autism Spectrum Disorders*. San Diego, CA: Elsevier/Academic Press.

McMurtry, S. (2013). *Parenting Style Differences in Black American and White American Young Adults*. Unpublished doctoral dissertation, University of Southern Mississippi, Hattiesburg, MS.

Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J. & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist* 51: 77-101.

Nihira, K. (1999). Adaptive behavior: A historical overview. In: R. Schalock & D. Braddock (eds.), *Adaptive Behavior and Its Measurement*, pp. 7-14. Washington DC: American Association on Mental Retardation.

Nihira, K., Foster, R., Shellhaas, M. & Leland, H. (1968). *Adaptive Behavior Scales Manual.* Washington, DC: American Association on Mental Deficiency.

Nihira, K., Leland, H. & Lambert, N. (1993). *Adaptive-Behavior Scale—Residential and Community: Examiner's Manual*, 2nd ed. Austin, TX: Pro-Ed.

Oakland, T.D. (2005). Adaptive behavior: Information critical for assessment and intervention. Invited workshop presented during the annual meeting of the International School Psychology Association, Athens, Greece.

Oakland, T.D. (2009). How universal are test development and use? In: E. Grigorenko (ed.), *Assessment of Abilities and Competencies in an Era of Globalization*, pp. 1-40. New York, NY: Springer.

Oakland, T.D. & Daley, M. (2013). Adaptive behavior: Its history, concepts, assessment, and applications. In: K.F. Geisinger (ed.), *APA Handbook of Testing and Assessment in Psychology*, vol. 3, pp. 183-212. Washington, DC: American Psychological Association.

Oakland, T.D. & Harrison, P.L. (2008). Adaptive behaviors and skills: An introduction. In: T. Oakland & P.L. Harrison (eds.), *Adaptive Behavior Assessment System-II: Clinical Use and Interpretation*, pp. 3-20. San Diego, CA: Elsevier/Academic Press.

Osborne, R.T. & McGurk, F.C.L. (1982). *The Testing of Negro Intelligence*, vol. 2. Athens, GA: Foundation for Human Understanding.

Pearson, R. (1992). *Shockley on Eugenics and Race*, pp. 15-49. Washington, DC: Scott-Townsend Publishers.

Peoples, C.E., Fagan, J.F. & Drotar, D. (1995). The influence of race on 3-year-old children's performance on the Stanford-Binet: Fourth Edition. *Intelligence* 21: 69-82.

Rushton, J.P. (1997). Race, intelligence, and the brain. *Personality and Individual Differences* 23: 169-180.

Rushton, J.P. (2000). *Race, Evolution, and Behavior: A Life History Perspective*, 3rd ed. Port Huron, MI: Charles Darwin Research Institute.

Rushton, J.P. (2001). Black–White differences on the g factor in South Africa: A "Jensen effect" on the Wechsler Intelligence Scale for Children—Revised. *Personality and Individual Differences* 31: 1227-1232.

Rushton, J.P. (2002). Jensen effects and African/Coloured/Indian/White differences on Raven's Standard Progressive Matrices in South Africa. *Personality and Individual Differences* 33: 1279-1284.

Rushton, J.P. & Jensen, A.R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy and Law* 11: 235-294.

Rushton, J.P. & Jensen, A.R. (2008). James Watson's most inconvenient truth: Race realism and the moralistic fallacy. *Medical Hypotheses* 71: 629-640.

Rust, J.O. & Wallace, M.A. (2004). Book Review: Adaptive Behavior Assessment System — Second Edition. *Journal of Psychoeducational Assessment* 22: 367-373.

Schalock, R.L. (1999). The merging of adaptive behavior and intelligence: Implications for the field of mental retardation. In: R.L. Schalock (ed.), *Adaptive Behavior and Its Measurement: Implications for the Field of Mental Retardation*, pp. 43-59. Washington, DC: American Association on Mental Retardation

Schalock, R.L. (2011). International perspectives on intellectual disability. In: K.D. Keith (ed.), *Cross-Cultural Psychology: Contemporary Themes and Perspectives*, pp. 312-328. New York, NY: Wiley-Blackwell.

Scheerenberger, R.C. (1983). *A History of Mental Retardation.* Baltimore: Paul H. Brookes.

Scott, R. & Sinclair, D. (1997). Ethnic-related cognitive profiles of black and white preschool children. *Journal of Comparative Human Biology* 28: 116-120.

Snyderman, M. & Rothman, S. (1990). *The IQ Controversy, the Media, and Public Policy*. Piscataway, New Jersey: Transaction.

Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. New York: Macmillan.

Spies, R.A. & Plake, B.S. (2005). *The Sixteenth Mental Measurements Yearbook*. Department of Educational Psychology at the University of Nebraska: Buros Institute of Mental Measurements.

Tasse, M., Schalock, R.L., Balboni, G., Bersani, H., Borthwick-Duffy, S., Spreat, S., Thissen, D., Widaman, K.F. & Zhang, D. (2012). The Construct of Adaptive Behavior: Its Conceptualization, Measurement, and Use in the Field of Intellectual Disability. *American Journal on Intellectual and Developmental Disabilities* 117(4): 291-303.

Tucker W.H. (2009). *The Cattell Controversy: Race, Science, and Ideology.* Champaign, IL: University of Illinois Press.

UNICEF (2016). *The State of the World's Children*. Retrieved 25 January 2017 from https://www.unicef.org/publications/files/UNICEF_SOWC_2016.pdf.

Wei, Y., Oakland, T. & Algina, J. (2008). Multigroup confirmatory factor analysis for the parent form, ages 5-21, of the Adaptive Behavior Assessment System-II. *American Journal on Mental Retardation* 113: 178-186.

Wilson, J. & Herrnstein, R.J. (1985). *Crime and Human Nature*. New York: Simon & Schuster.

Woolf, S., Woolf, C.M. & Oakland, T.D. (2010). Adaptive behavior among adults with intellectual disabilities and its relationship to community independence. *Intellectual and Developmental Disabilities* 48: 209-215.

Yeargin-Allsopp, M., Drews, C., Decoufle, P. & Murphy, C.C. (1995). Mild mental retardation in Black and White children in metropolitan Atlanta: A case control study. *American Journal of Public Health* 85: 324-328.

MANKIND QUARTERLY 2017 57:4 659-668

Differences in the Intelligence of 15 Year Olds in 42 Provinces and Cities of the Russian Federation and Their Economic, Social and Geographical Correlates

Richard Lynn*
University of Ulster, UK

Helen Cheng
University College London, UK

Andrei Grigoriev
Institute of Psychology, Russian Academy of Sciences, Moscow, Russia

*Corresponding author: lynnr540@aol.com

This study reports data for the intelligence of 15 year olds in 42 provinces and cities of the Russian Federation assessed in the 2015 Programme for International Student Assessment (PISA) and their economic, social and geographical correlates. It was found that the average PISA scores of the provinces and cities were significantly correlated with the scores on the Unified State Examination in 2014 (r = .53, p<.001) reported by Grigoriev et al. (2016), with literacy rates in 1897 (r = .50, p<.01), with the percentage of ethnic Russians in the population, and with latitude and longitude showing that PISA scores were higher in the more northerly and westerly provinces.
Key Words: PISA 2015; Russia; Russian regions; Per capita income; Latitude; Longitude; Ethnicity.

There have been a number of studies of regional differences in intelligence within countries and their social, economic, demographic and epidemiological correlates. These include positive associations with per capita income, educational attainment, life expectancy and stature, and negative associations with infant mortality and fertility. The first of these studies was reported for the United States by Davenport and Remmers (1950), who gave a correlation of .32 between state IQ and per capita income. Later studies reported similar associations for the regions of the British Isles (Lynn,1979), France (Lynn, 1980; Kirkegaard, 2015), the United States (McDaniel, 2006; Shatz, 2009), Italy (Lynn, 2010; Piffer & Lynn, 2014), Portugal (Almeida, Lemos & Lynn, 2011), Germany (Roivainen, 2012), Spain (Lynn, 2012), China (Lynn & Cheng, 2013; Lynn, Cheng & Wang, 2016), Japan (Kura, 2013), Finland (Dutton & Lynn, 2014), India (Lynn & Yadav, 2015), Turkey (Lynn, Sakar & Cheng, 2015), the United Kingdom (Carl, 2015) and Brazil (Lynn, Antonelli-Ponti, Mazzei & Da Silva, 2017).

There have been two studies of regional differences in intelligence and their economic and social correlates in Russia. In the first of these, Grigoriev, Lapteva and Lynn (2009) reported data for European Russia in the late nineteenth century. Differences in intelligence for 50 provinces were inferred from the literacy rates and were shown to be significantly negatively associated with infant mortality (r = -.28), fertility (r = -.75) and longitude (r = -.43) showing that IQs were higher in the more westerly provinces, and significantly positively associated with stature (r = .56) and latitude (r = .33) showing that IQs were higher in the more northerly provinces.

In the second study, Grigoriev et al. (2016) reported data for differences in educational attainment (educational quotient, EQ) as a proxy for intelligence, a number of socio-economic variables, and latitude and longitude for 79 provinces of the Russian Federation. The average intelligence of the provinces was significantly positively correlated with urbanization (r = .43), the percentage of ethnic Russians (r = .39), net migration (r = .54) and latitude (r = .35), such that intelligence was higher in the North, and significantly negatively correlated with infant mortality (r = -.43), fertility (r = -.39) and longitude (r = -.36), such that intelligence was higher in the West. In the present paper we report a further study of differences in intelligence across provinces and cities in the Russian Federation and their social and economic correlates.

Method

There are 83 provinces in the Russian Federation. Data for 40 provinces of these and two cities (Moscow City and St. Petersburg City) are given for 6,036 students aged 15 who participated in the Programme for International Student Assessment (PISA) in 2015 in tests of Reading, Math, and Science. The PISA scores have been adopted as measures of regional IQs in studies in Turkey (Lynn, Sakar & Cheng, 2015) and Brazil (Lynn, Antonelli-Ponti, Mazzei & Da Silva (2017). This usage is justified because the PISA tests are constructed to measure cognitive ability rather than knowledge of the curriculum and by the finding that across 108 national populations PISA scores are correlated with IQs at *r* = .91 (Lynn & Meisenberg, 2010).

Differences in intelligence reported for the Russian provinces by Grigoriev et al. (2016) are included in the analyses to ascertain how far they are consistent with the PISA scores. We also examined the relationship of the PISA scores with contemporary educational attainment in the regions of the Russian Federation, and literacy rates of the corresponding regions of the former Russian Empire in the late nineteenth century. The administrative map of the contemporary Russian Federation differs from that of the former Russian Empire. To ensure an approximate correspondence between the two sets of regions, we used the following procedure: The provinces of the Russian Empire were divided into smaller divisions (*uyezds*). Data for literacy in the *uyezds* are available in the Russian 1897 census (Первая всеобщая перепись населения Российской Империи 1897, 1899-1905). We assume that if the main town of a given *uyezd* is on the territory of a contemporary region, then the whole *uyezd* is in this region. Thus the contemporary regions are regarded as consisting of former *uyezds*, which may belong to several former provinces. The literacy rates of populations in the territories of the contemporary regions in the late nineteenth century were calculated as weighted means of the rates for former *uyezds* whose main towns are in the territories of the contemporary regions. There were no main towns in the former *uyezds* in the territory of Yamalo-Nenets autonomous district, and the literacy rate for this district was estimated from the literacy rate for the *uyezd* in Tobolsk province that is now the greatest part of this district.

The data for per capita income were measured as GDP (gross domestic product) per capita based on the UN figures in the year 2009 (UNDP, 2011). The percentage of Russian ethnicity was taken from the 2010 All-Russian Population Census published by the Russian Federal State Statistics Service (2011).

Results

Table 1 gives the means and standard deviations of the PISA total scores, the educational attainment quotients (EQ) given by Grigoriev et al. (2016), literacy rates in 1897, GDP per capita, latitude, longitude, and the percentage of ethnic Russians. The PISA scores on Reading, Math and Science were highly intercorrelated: Reading x Math: 0.83; Reading x Science: 0.91; Math x Science: 0.92. The scores are therefore averaged for the correlations shown in Table 2.

A multiple regression analysis using the PISA scores as the dependent variable and the other measures as the independent variables showed that latitude (β = .34, t = 2.29, p<.05), longitude (β = -.29, t = 2.23, p<.05) and the percentage of Russian ethnicity (β = .36, t = 2.68, p<.01) were the significant predictors of the PISA scores, accounting for 41% of total variance.

Table 1. *Provincial and city differences in PISA 2015, the state examination scores (EQ) in 2014, the retrieved literacy rates in 1897, and other variables in Russia.*[1]

Region	PISA Score Mean ± SD	EQ	Lit	GDP	% Russian	Lat.	Long.	N
Rep. of Bashkortostan	469.0 ± 75.9	99	16.0	15,797	36.1	54.3	56.2	183
Rep. of Dagestan	424.1 ± 71.0	84	10.6	9,337	3.6	43.1	46.5	135
Kabardino-Balkarian Rep.	454.0 ± 59.4	82	5.0	7,666	22.5	43.4	43.2	111
Rep. of Sakha (Yakutia)	469.1 ± 65.0	96	4.1	21,159	37.8	66.2	129.1	96
Rep. of Tatarstan	487.5 ± 65.4	104	18.9	23,290	39.7	55.5	49.1	129
Chuvashi Rep.	505.0 ± 60.5	101	12.3	10,971	26.9	55.3	47.1	93
Altai territory	493.1 ± 78.2	101	8.8	7,520	94.0	52.5	83.8	98
Krasnodar territory	467.8 ± 71.3	105	18.1	13,899	88.3	45.2	39.1	173
Krasnoyarsk territory	496.9 ± 78.2	98	13.9	20,779	91.3	59.6	83.2	168
Primorsky territory	491.4 ± 74.9	96	29.5	12,574	92.5	45.2	131.9	123
Stavropol territory	454.3 ± 57.0	97	16.4	8,725	80.9	45.1	41.9	130
Arkhangelsk region	501.4 ± 64.0	103	21.6	19,310	95.6	63.3	40.3	129
Belgorod region	482.2 ± 68.1	100	15.9	19,569	94.4	50.5	36.7	120
Vladimir region	469.5 ± 70.8	98	26.4	11,666	95.6	56.1	40.4	142
Volgograd region	471.8 ± 68.1	96	25.5	13,200	90.0	48.4	44.5	103
Voronezh region	500.0 ± 69.0	100	16.6	11,036	95.5	51.1	40.1	81
Ivanovo region	509.1 ± 64.5	97	26.8	7,425	95.6	57.1	41.1	154
Irkutsk region	452.8 ± 74.6	93	15.2	15,987	88.5	57.2	106.0	175
Kaliningrad region	518.8 ± 61.8	103	-	14,136	86.4	54.5	21.1	150
Kamchatka territory	482.2 ± 62.0	86	12.1	12,931	85.9	56.0	159.0	109
Kemerovo region	460.4 ± 60.1	99	8.8	18,721	93.7	54.6	87.1	103
Kostroma region	522.1 ± 66.6	96	26.7	10,941	96.6	58.3	43.4	115

Region	PISA Score Mean ± SD	EQ	Lit	GDP	% Russian	Lat.	Long.	N
Lipetzk region	523.7 ± 68.2	93	16.7	17,902	96.3	52.4	39.6	86
Moscow region	511.3 ± 71.4	110	27.0	17,255	91.6	55.5	38.2	286
Nizhni Novgorod region	502.2 ± 64.8	104	20.3	14,709	95.1	56.3	44.3	169
Novgorod region	476.7 ± 68.2	97	23.6	16,397	95.1	58.3	32.2	89
Novosibirsk region	485.3 ± 74.8	106	9.4	13,383	93.1	55.3	79.3	112
Omsk region	493.9 ± 65.7	98	12.3	16,213	85.8	56.1	73.1	120
Orenburg region	503.1 ± 71.7	100	18.0	19,507	75.9	51.5	55.1	86
Perm territory	498.6 ± 81.0	105	18.5	16,642	87.1	59.1	56.3	189
Rostov region	464.4 ± 76.7	98	23.0	11,302	90.3	47.5	41.1	206
Ryazan region	496.6 ± 68.7	102	18.8	11,510	95.1	54.2	40.4	113
Samara region	488.8 ± 69.5	101	19.8	14,520	85.6	53.3	50.3	191
Saratov region	509.5 ± 72.9	99	26.1	12,812	90.0	51.5	46.4	109
Sverdlovsk region	495.4 ± 75.4	106	20.6	15,811	90.6	58.4	61.2	232
Tomsk region	480.6 ± 72.6	108	19.3	19,064	92.1	58.5	82.1	79
Ulyanovsk region	500.5 ± 62.4	97	17.6	11,794	73.6	53.5	47.6	116
Chelyabinsk region	500.7 ± 69.9	98	19.2	15,098	83.8	54.3	60.2	206
Moscow City	516.4 ± 72.8	110	53.1	40,805	91.7	55.5	38.2	373
St. Petersburg City	524.4 ± 71.1	111	61.6	25,277	84.7	59.6	30.3	245
Khanty-Mansijsk autonomous district - Yugra	480.3 ± 64.7	94	8.6	-	98.1	62.2	70.1	119
Yamalo-Nenets autonomous district	481.6 ± 75.4	-	8.4	-	61.7	66.7	66.5	90
Russian Federation	**491.8 ± 73.2**	**99.1**	**19.3**	**18,869**	**81.0**	**51.5**	**58.7**	**6,036**

[1] Observations (N) were unweighted, means were weighted with Senate Weight (country weight). EQ = Educational quotients, the Unified State Examination in 2014 from Grigoriev et al. (2016). Lit = % Literacy 1897, from Тройницкий (1899-1905). GDP = per capita gross domestic product, from UNDP (2011), based on the UN figures in the year 2009. %Russian = Russian ethnicity, from Russian Federal State Statistics Service (2011). Lat = latitude; Long = longitude.

Table 2. *Mean ± standard deviation and correlation matrix for variables shown in Table 1. *p<.05; **p<.01; ***p<.001.*

	Mean ± SD	Lat.	Long.	PISA	EQ	Lit.	GDP
Longitude	58.65 ± 29.18	.16					
PISA	491.8 ± 73.2	.35*	-.25				
EQ	99.05 ± 6.69	.36*	-.31	.53***			
Literacy	0.19 ± 0.11	.02	-.37*	.50**	.52**		
GDP	4.17 ± 0.15	.47**	.07	.31	.55***	.41**	
%Russian	79.08 ± 23.47	.11	-.04	.45**	.43**	.36*	.17

Discussion

There are six points of interest in the study. First, the PISA scores were significantly correlated (r = .53, p<.001) with the educational quotients (EQs) of the provinces estimated from educational achievement tests given by Grigoriev et al. (2016) and with the literacy rates in 1897 (r = .50, p<.01) showing a significant degree of consistency between the three measures. This consistency is remarkable considering the smallish sample sizes for PISA in individual provinces and the limited range of variation among the provinces. Note in particular that the province of Dagestan has the lowest PISA score (424.1) and the second lowest EQ (84); and also that the city of St. Petersburg has the highest PISA score (524.4), the highest EQ (111) and the highest literacy rate in 1897 (61.6%). The city of Moscow has the fourth highest PISA score (516.4), the second highest EQ (110) and the second highest literacy rate in 1897 (53.1%). These high results for St. Petersburg (the former capital city) and Moscow (the present capital city) are consistent with several studies reporting that the populations of capital cities have higher IQs than the rest of the population, e.g. in the British Isles and France (Lynn, 1979; 1980) and in Portugal (Almeida et al., 2011). The likely explanation for this is that there is a tendency for those with higher than average intelligence to migrate from the provinces to capital cities, establish families there and transmit their high intelligence to later generations.

Second, the PISA scores were correlated at r = .31 with GDP per capita. The correlation falls just short of statistical significance at p<.05 (r = .32 would be statistically significant). GDP per capita was also significantly correlated at .55 (p<.001) with the EQs of the provinces and cities given by Grigoriev et al. (2016) and with the literacy rates in 1897 (r = .41, p<.01) providing further confirmation of this association. Taken together, these results are consistent with a number of previous studies reporting positive associations of regional IQs with per capita income cited in the introduction.

Third, the PISA scores were significantly correlated at $r = .45$ ($p<.01$) with the percentage of the population with Russian ethnicity. This result is confirmed by the multiple regression analysis showing that the percentage of Russian ethnicity was a significant predictor of the PISA scores ($\beta = .36$, $t = 2.68$, $p<.01$). This result is closely similar to the .39 EQ correlation obtained by Grigoriev et al. (2016). Note in particular that the province of Dagestan has the lowest percentage with Russian ethnicity (3.6 percent), the lowest PISA score (424.1) and the second lowest EQ (84). Studies of the intelligence of the ethnic minorities in Russia have reported varying results. In the province of Yakutia the Yakuts obtained approximately the same IQ as the ethnic Russians with British IQs of 97.0 and 97.9, respectively (Shibaev & Lynn, 2017). However, a sample of ethnic Evenk/Tungus of the Russian Far East obtained a British IQ of only 80 (Shibaev & Lynn, 2015).

Fourth, the PISA scores were significantly correlated at $r = .35$ ($p<.05$) with latitude showing that IQs are higher in the more northerly provinces. This result is confirmed by the multiple regression analysis showing that latitude was a significant predictor of the PISA scores ($\beta = .34$, $t = 2.29$, $p<.05$). This result confirms that obtained in the late nineteenth century in which intelligence was found to be higher in the North than in the South ($r = .33$, $p<.05$) reported by Grigoriev, Lapteva and Lynn (2009). These correlations are virtually the same as the .35 correlation obtained for 79 Russian provinces reported by Grigoriev et al. (2016) and confirmed at .36 in Table 2 for the 40 provinces and two cities in the present study. Note in particular that the most southerly province of Dagestan has the lowest PISA score (424.1) and the second lowest EQ (84). These results corroborate the 'cold winters theory' of the evolution of racial differences in intelligence advanced in Lynn (1991, 2015) that proposes that greater cognitive demands for survival in the temperate and cold environments of the northern hemisphere provided a selection pressure for the evolution of higher intelligence. This theory has been corroborated by a number of studies that have reported statistically significant associations between average cold winter temperatures and national IQ (Kanazawa, 2008; Meisenberg & Woodley, 2013; Rushton, 2000; Templer & Arikawa, 2006).

Fifth, the PISA scores were negatively correlated at $r = -.25$ with longitude showing that IQs tended to be higher in the more westerly provinces. This correlation is not statistically significant but the multiple regression analysis showed that longitude was a significant predictor of the PISA scores ($\beta = -.29$, $t = 2.23$, $p<.05$). This result is in the same direction as the negative correlation of longitude with EQ (-.31, ns) and with literacy in 1897 ($r = -.37$, $p<.015$). Note in particular that the most westerly province of Kaliningrad has the third highest PISA score (518.8) and the tenth highest EQ (103). These results taken together show that there is a significant tendency for IQs to be higher in the West of the Russian Federation.

Sixth, the PISA results give scores for one province for which no data are given by Grigoriev et al. (2016). This is the north central province of the Yamalo-Nenets autonomous district which has a below average score of 481.6. Part of the explanation for this low score may be that the percentage of ethnic Russians is relatively low at 61.7 as compared with 79 percent for the whole of the Russian Federation. Another part of the explanation may be its remote Arctic location.

References

Almeida, L.S., Lemos, G.C. & Lynn, R. (2011). Regional differences in intelligence and per capita incomes in Portugal. *Mankind Quarterly* 52: 213-221.

Carl, N. (2016). IQ and socio-economic development across regions of the UK. *Journal of Biosocial Science* 48: 406-417.

Davenport, K.S. & Remmers, H.H. (1950). Factors in state characteristics related to average A-12 V-12 scores. *Journal of Educational Psychology* 41: 110-115.

Dutton, E. & Lynn, R. (2014). Regional differences in intelligence and their social and economic correlates in Finland. *Mankind Quarterly* 54: 447-456.

Grigoriev, A. Lapteva, E. & Lynn, R. (2009). Studies of socio-economic and ethnic differences in intelligence in the former Soviet Union in the early twentieth century. *Intelligence* 37: 447-452.

Grigoriev, A., Ushakov, D., Valueva, E., Zirenko, M. & Lynn, R. (2016). Differences in educational attainment, socio-economic variables and geographical location across 79 provinces of the Russian Federation. *Intelligence* 58: 14-17.

Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence* 36: 99-108.

Kirkegaard, E.O.W. (2015). IQ and socio-economic variables in the French departments: Reanalysis and new data. *Mankind Quarterly* 56: 113-135.

Kura, K. (2013). Japanese north–south gradient in IQ predicts differences in stature, skin color, income, and homicide rate. *Intelligence* 41: 512-516.

Lynn, R. (1979). The social ecology of intelligence in the British Isles. *British Journal of*

Social & Clinical Psychology 18: 1-12.

Lynn, R. (1980). The social ecology of intelligence in France*. British Journal of Social & Clinical Psychology* 19: 325-331*.*

Lynn, R. (1991). The evolution of race differences in intelligence. *Mankind Quarterly* 32: 99-173.

Lynn, R. (2010). In Italy, north–south differences in IQ predict differences in income, education, infant mortality, stature, and literacy. *Intelligence* 38: 93-100.

Lynn, R. (2012). North-south differences in Spain in IQ, educational attainment, per capita income, literacy, life expectancy and employment. *Mankind Quarterly* 52: 265-291.

Lynn, R. (2015). *Race Differences in Intelligence: An Evolutionary Analysis*. Augusta, GA: Washington Summit.

Lynn, R. & Cheng, H. (2013). Differences in intelligence across thirty-one regions of China and their economic and demographic correlates. *Intelligence* 41: 553-559.

Lynn, R. & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence* 38: 353-360.

Lynn, R. & Yadav, P. (2015). Differences in cognitive ability, per capita income, infant mortality, fertility and latitude across the states of India. *Intelligence* 49: 179-185.

Lynn, R., Sakar, C. & Cheng, H. (2015). Regional differences in intelligence, income and other socio-economic variables in Turkey. *Intelligence* 50: 144-150.

Lynn, R., Cheng, H. & Wang, M.-R. (2016). Differences in the intelligence of children across thirty-one provinces and municipalities of China and their economic and social correlates. *Intelligence* 58: 10-13.

Lynn, R., Antonelli-Ponti, M., Mazzei, R.F., Da Silva, J.A. & Meisenberg, G. (2017). Differences in intelligence, income, health, life expectancy, fertility and homicide across the twenty seven states of Brazil. *Mankind Quarterly*, this issue.

McDaniel, M.A. (2006). State preferences for the ACT versus SAT complicates inferences about SAT-derived state IQ estimates: A comment on Kanazawa. *Intelligence* 34: 601-606.

Meisenberg, G. & Woodley, M.A. (2013). Global behavioral variation: A test of differential-K. *Personality and Individual Differences* 55: 273-278.

Piffer, D. & Lynn, R. (2014). New evidence for differences in fluid intelligence between north and south Italy and against school resources as an explanation for the north–south IQ differential. *Intelligence* 46: 246-249.

Roivainen, E. (2012). Economic, educational and IQ gains in eastern Germany 1990-

2006. *Intelligence* 40: 571-575.

Rushton, J.P. (2000). *Race, Evolution and Behavior*. Port Huron, MI: Charles Darwin Research Institute.

Russian Federal State Statistics Service (2011). *2010 All-Russian Population Census, vol. 1.*

Shatz, S.M. (2009). State IQ and fertility in the United States. *Mankind Quarterly* 49: 38-49.

Shibaev, V. & Lynn, R. (2015). The intelligence of the Evenk/Tungus of the Russian Far East. *Mankind Quarterly* 56: 202-207.

Shibaev, V. & Lynn, R. (2017). The intelligence of Yakuts and ethnic Russians in Yakutia. *Mankind Quarterly*, this issue.

Templer, D.I. & Arikawa, H. (2006). Temperature, skin color, per capita income and IQ: An international perspective. *Intelligence* 34: 121-139.

Тройницкий Н.А. (ed.). (1899-1905). Первая всеобщая перепись населения Российской Империи 1897 г. Retrieved from http://forum.relicvia.ru/topic/14896-pervaia-vseobschaia-perepis-naseleniia-1897-goda/?hl=%2Bвсеобщая+%2Bперепись.

UNDP (2011). *National Human Development Report for the Russian Federation 2011*.

MANKIND QUARTERLY 2017 57:4 669-679

Intelligence and Intellectual Property Rights Protection: The Global Evidence

Shoirahon Odilova,* Gu Xiaomin
Donghua University, China

*Address for correspondence: Glorious Sun School of Business & Management, Donghua University, 1882 West Yan'an Road, Shanghai, China 200051; email: shoiraodilova@yahoo.com

This study explores the relationship between the average intelligence in countries and the protection of intellectual property rights, using data for 127 developing and developed countries for the year 2016. The results provide significant support for the conjecture that intelligence is positively associated with intellectual property protection. For example, a 10 points increase in national IQ leads to a 0.8 points increase in the Intellectual Property Rights (IPR) index, slightly more than half a standard deviation.
Key Words: Intelligence, Intellectual property protection, IQ.

A large strand of empirical literature explores the effect of intelligence on economic growth and economic development as measured by GDP per capita and its proxies (Salahodjaev, 2015a; Whetzel & McDaniel, 2006; Ram, 2007). Other studies show that intelligence is positively correlated with correlates of economic growth such as rule of law (Kanyama, 2014) and financial development (Kodila-Tedika & Asongu, 2015), and negatively with corruption (Potrafke, 2012).

The extant literature indicates that intelligence is a robust predictor of the quality of legal frameworks, but the association between intelligence and intellectual property rights (IPR) protection has not yet been investigated. The relationship is interesting because IPR protection is positively associated with economic growth (Thompson & Rushing, 1996). This raises the possibility that intelligence may have indirect effects on economic growth via stronger IPR enforcement. This possibility is explored in the present study. There are a number of channels through which intelligence may be linked to IPR protection: (1) economic development, which provides the financial resources for IPR enforcement; (2) quality of institutional arrangements; and (3) the human capital (aka competence) of the individuals who are in charge of IPR enforcement.

Published studies suggest that human capital, as proxied by national IQ, is a successful predictor of economic development. For example, Whetzel and McDaniel (2006) revisit the association between intelligence and GDP per capita after controlling for democracy, economic freedom and oil production. They document that intelligence correlates significantly and positively with GDP per capita. In a follow-up study, Hunt and Wittman (2008) further find that intelligence is significantly related to economic development, although they 'question the simple explanation that national intelligence causes national wealth.' Similar findings were also obtained by Dickerson (2006) and Kanazawa (2006).

Apart from economic development, intelligence may also predict quality of IPR laws because it generally predicts the quality of institutions. Indeed, Potrafke (2012) finds that high-IQ individuals have longer time horizons and thus may break intellectual property rights less. Similarly, Salahodjaev (2015c), using data from 158 countries over the period 1999–2007, documents that national IQs are negatively and significantly related to the size of the shadow economy. Moreover, countries with higher levels of cognitive abilities are associated with stronger rule of law (Kanyama, 2014) and quality of institutional arrangements (Meisenberg, 2012, 2014). In this study, we contribute to the emerging strand of interdisciplinary research on the consequences of intelligence by investigating the relationship between intelligence and IPR protection in a cross-section of 127 countries.

Our findings suggest that countries with higher average IQ (or higher IQ of the educated classes) have more stringent regulations to ensure IPR protection. This association is highly statistically significant ($p<.01$ in the majority of our models) even when we control for GDP per capita, ethnic diversity and a dummy variable for countries with British civil law. In particular, when cognitive abilities increase by 10 IQ points, the IPR index increases by 0.8 points, slightly more than half a standard deviation.

Data & Methods

Dependent variable

In this study we use two proxies for intellectual property rights (IPR) protection, namely, the IPR index from Property Rights Alliance, and the Intellectual Property Protection (IPP) index from World Economic Forum.

Intellectual Property Rights (IPR) Index

The International Property Rights Index (IPRI) was developed by Property Rights Alliance to function as a barometer for the status of property rights across the world. The IPRI consist of three core sub-dimensions: Legal and Political Environment (LP), Physical Property Rights (PPR), and Intellectual Property

Rights (IPR). The LP component assesses the strength of the institutions of a country, the respect for the 'rules of the game' among citizens; consequently, the measures used for the LP are broad in scope. This component has a significant impact on the development and protection of physical and intellectual property rights. The other two components of the index, PPR and IPR, reflect two forms of property rights that are crucial to the economic development of a country. The items included in these two categories account for both de jure rights and de facto outcomes of the countries considered.[1] The overall grading scale of the IPRI ranges from 0 to 10, where 10 is the highest value for a property rights system and 0 is the lowest value (i.e. most negative) for a property rights system within a country.

Intellectual Property Protection (IPP) Index

The IPP index is a component of the Global Competitiveness Index of the World Economic Forum, obtained from the 2015-2016 Global Competitiveness Report (http://reports.weforum.org/global-competitiveness-report-2015-2016/). It is based on opinion surveys reflecting a nation's protection of intellectual property. Expert participants in each country were asked to rate their nation's intellectual property protection, scoring it from "weak and not enforced" (1) to "strong and enforced" (7). The correlation between IPR and IPP index is $r = .882$ (N = 128 countries). This is higher than the usual correlations among development indicators, and shows that the two indices measure the same or nearly the same construct.

National IQs

As a measure of intelligence we use the dataset compiled by Lynn and Vanhanen (2012). Although some scholars attempted to disprove the importance of national IQs in cross-country studies (Barnett & Williams, 2004; Volken, 2003), ample studies have been published over the past decade proving that intelligence is significantly associated with numerous socio-economic outcomes (Kanyama, 2014; Potrafke, 2012; Salahodjaev, 2015b, 2016).

Lynn and Vanhanen (2012) compiled studies across the world in which IQ tests had been administered. Based on this evidence, the authors estimated each country's average IQ score by setting the intelligence quotient of the UK at 100 ± 15, and scaling the IQ scores of other countries to this metric ("Greenwich IQ").

[1] See http://internationalpropertyrightsindex.org/ipri2016_comp for more details

To graphically illustrate the link between national IQ and intellectual property protection, we present the scatterplot between national IQs and IPR index. Figure 1 shows that overall cognitive abilities at the national level are positively related to the IPR index. The correlation between intelligence and IPR index is *r* = .63. This preliminary result suggests that there is a close link between IQ and intellectual property rights. In fact, intellectual property rights protection is closely related to the quality of institutions in general, as indexed by the Governance Indicators of the World Bank, especially government effectiveness, regulatory quality, rule of law, and control of corruption. The correlation between IPR index and market freedoms is .65 (Ginarte & Park, 1997). We can even hypothesize that intellectual property rights may be a link between IQ and economic growth (Meisenberg, 2014). Moreover, Odilova and Gu (2016) shows that there is a joint effect of intelligence and IPR protection on economic growth. This study suggests the possibility that intellectual property rights may *impair* economic growth, perhaps because most of the time they benefit only a few big companies and prevent the widespread adoption of innovations. However, the study also reports that cognitive abilities may offset this negative effect and in high-IQ countries IPR protection is a growth-enhancing variable.

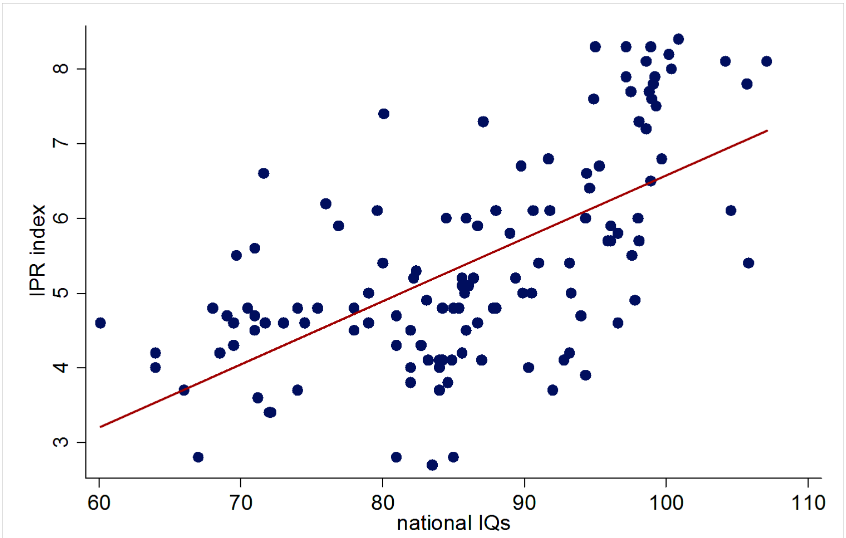

Figure 1a. *Scatter plot between national IQs and IPR index; Correlation coefficient r = .63; N = 125 countries.*

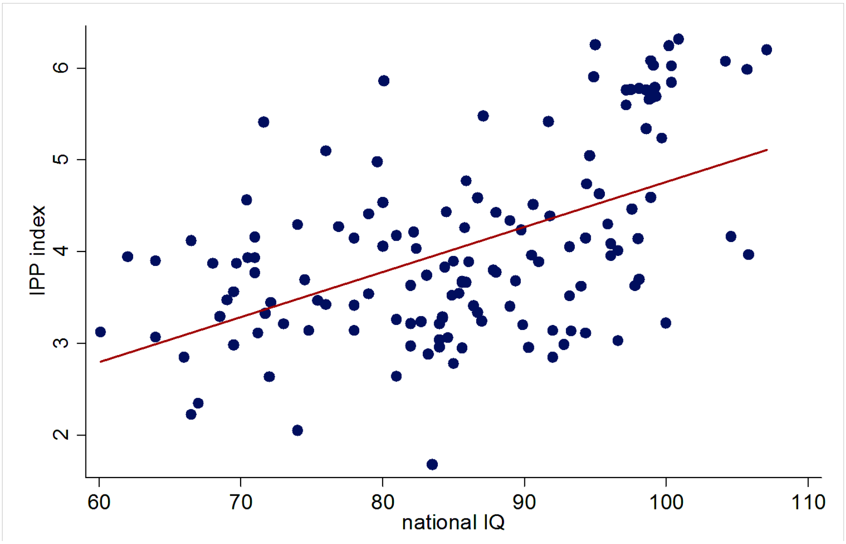

Figure 1b. *Scatterplot between national IQs and IPP index; Correlation coefficient r = .52; N = 139 countries.*

Results

Although Figure 1 suggests that national IQs and IPI index are positively associated, the meaning of this relationship is not immediately obvious. Therefore to explore the link between intelligence and IPR index we estimate an econometric model of the general form:

$$IPRI = CONSTANT + b * IQ + c * X$$

where IPRI is one the measures of intellectual property rights protection; IQ is mean national IQ; X is the vector of control variables such as GDP per capita, ethnic diversity, and a dummy variable for British civil law.

The statistical modeling results are presented in Tables 1 and 2. Table 1 reports the results for the IPRI index from Property Rights Alliance. Model 1 presents simple bivariate regression between IPR index and national IQ. The coefficient for IQ is positive and statistically significant at the 1% level. For example, when cognitive abilities increase by 10 points, IPR index increases by 0.8 points, slightly more than half a standard deviation. On the other hand, it is documented that cognitive abilities are positively correlated with economic development at the country level. Therefore, the estimates in Model 1 may capture the effect of GDP per capita on intellectual property protection, and also a reverse effect of intellectual property protection on GDP per capita. Thus, in Model 2 we now include GDP per capita for 2012 (the same year as national IQ) from World Bank. The coefficient of GDP per capita is positive and significant at the 1% level. The IQ estimate is still statistically significant, although its magnitude decreases by nearly half.

In Models 3 and 4 we include the ethnic diversity index from Alesina et al. (2003) and a dummy variable for countries with British civil law to capture the effect of quality of legal origins and culture on IPR protection. The results suggest that countries with British civil law enjoy stronger protection of intellectual property rights. However, the estimate for ethnic diversity is not significant. The estimate for national IQ is positive and statistically significant at the 1% level. This specification explains more than 65% of cross-national variations in IPR protection.

Table 1. *Regression models with IPRI as dependent variable. B coefficients are shown with standard errors in parentheses.* * *p<0.1,* ** *p<0.05,* *** *p<0.01.*

	(1)	**(2)**	**(3)**	**(4)**
IQ	0.0843***	0.0441***	0.0422***	0.0475***
	(0.0093)	(0.0086)	(0.0104)	(0.0101)
GDP per capita		0.0385***	0.0385***	0.0366***
		(0.0045)	(0.0046)	(0.0044)
Ethnic diversity			-0.1379	-0.2498
			(0.3771)	(0.3637)
British civil law				0.5718***
				(0.1715)
Constant	-1.8550**	0.8523	1.0850	0.5459
	(0.8056)	(0.7036)	(0.9533)	(0.9296)
N	127	125	124	124
adj. R^2	0.3944	0.6280	0.6257	0.6548

We repeat this exercise for the IPP index from the World Economic Forum. Table 2 shows again a strong positive link between intelligence and intellectual property protection, which remains significant at the 1% level in this sample of 136 nations. One possible link through which intelligence may strengthen intellectual property protection is by the rule of law and decreased corruption (Kanyama, 2014; Potrafke, 2012).

Table 2. *Regression models with IPP as dependent variable. B coefficients are shown with standard errors in parentheses.* * *p<0.1,* ** *p<0.05,* *** *p<0.01.*

	(1)	(2)	(3)	(4)
IQ	0.0491***	0.0188***	0.0172**	0.0237***
	(0.0068)	(0.0066)	(0.0079)	(0.0076)
GDP per capita		0.0303***	0.0304***	0.0282***
		(0.0036)	(0.0037)	(0.0035)
Ethnic diversity			-0.1181	-0.2083
			(0.2992)	(0.2827)
British civil law				0.5577***
				(0.1323)
Constant	-0.1477	1.8734***	2.0649***	1.4281**
	(0.5884)	(0.5340)	(0.7257)	(0.7001)
N	139	137	136	136
adj. R^2	0.2695	0.5237	0.5205	0.5745

Finally we checked whether the link between intelligence and IPR protection varies with the level of economic development. To do so we regressed our dependent variables on standardized national IQ (zIQ), standardized log-transformed GDP per capita (zlnGDP) and their cross-product. We find that the coefficient for IQ remains positive although it no longer reaches a conventional statistical significance level of .05, and the interaction term is strongly positive and significant. This implies that intelligence has a strong positive effect on IPR protection in high income countries but not in low-income countries. Variables that are statistically significant at p<.05 or below are denoted in bold in the regression equation:

$$\text{IPRI} = 4.729 + 1.075\,\mathbf{zlnGDP} + 0.092\,\text{zIQ} + 0.638\,\mathbf{zIQ} * \mathbf{zlnGDP}$$

This is also seen when the sample is divided into those with high per capita GDP and those with low per capita GDP. A regression analysis was performed separately for the two kinds of country, with IQ and logged GDP per capita as predictors. It shows that in the group of poor countries, IQ does not predict IPRI independently of GDP, but it does have an independent effect in the sample of rich countries (Table 3). One possible reason is that rich countries are the beneficiaries of intellectual property protection because intellectual property rights are held mainly by individuals and companies in rich countries.[2] Therefore, protection of intellectual property rights is in the self-interest of high-income countries but not necessarily of low-income countries who depend on products that are protected by patents or copyright in high-income countries.

Table 3. *Prediction of intellectual property rights index (IPRI) in subsamples of high-income ("rich") and low-income ("poor") countries. B coefficients are shown with standard errors in parentheses.* * *p<0.1,* ** *p<0.05,* *** *p<0.01.*

	poor	**rich**
IQ	0.0200	0.0916***
	(0.0143)	(0.0167)
GDP per capita (log)	0.5038***	0.7424*
	(0.1421)	(0.4219)
Constant	2.2638**	-4.1755
	(0.9817)	(2.6258)
N	104	21
adj. R^2	0.3816	0.5835

Conclusion

The results of this study have a number of important implications. First, this study further shows that cognitive ability is a robust predictor of one of the dimensions of quality of institutions, namely, intellectual property rights (IPR) protection. Earlier studies explored the effect of national IQ on rule of law, corruption perceptions, shadow economy and economic freedoms overlooking the importance of IPR protection.

Second, the findings in this study may further suggest that IPR protection has indirect and positive effects on economic growth in cognitively able economies. Moreover, as suggested by Kanyama (2014, p. 44), enforcing protection of intellectual property 'require[s] the ability to understand the principles that rule them in order to produce a high quality outcome' such as innovation and increasing productivity and efficiency. Thus, our findings further underline the importance of human capital not only in economic growth but also in managing institutions. High-IQ nations tend to have legal systems that protect all kinds of things, including intellectual property and economic freedom, promoting entrepreneurship that further increases welfare.

2 We thank a referee for this suggestion

Finally, the results reported in Table 3 are best explained if we assume that high-income countries but not low-income countries are beneficiaries of IPR protection. Therefore, for less developed countries strengthening intellectual property rights may be harmful for economic development as suggested in Adams (2009).

However, our study has a number of limitations. First, there is ongoing criticism of the national IQ dataset compiled by Lynn and co-authors. For example, Hunt (2011, p. 439) has argued that "Lynn and Vanhanen disregard any question about the validity of various intelligence tests across different countries and cultures." Moreover, Nechyba (2004, p. 1178) has highlighted the "relatively weak statistical evidence and dubious presumptions [on the link between intelligence and wealth of nations]." However, in their follow up studies, Lynn and Vanhanen (2012) and others have shown that intelligence is highly correlated with results on student achievement tests and various socio-economic outcomes. Moreover, Lynn and Vanhanen (2012) further show that in countries with multiple IQ tests administered the correlations for these scores are more than .9.

This study explores the association between intelligence and soft measures of IPR protection, as measured by indices. More specific associations between intelligence and the size of the illegal copyright market or the damage to the economy from software piracy remains an avenue for future research. Future studies should also explore the perceptions of software piracy and intelligence at the micro (individual) level.

References

Adams, S. (2009). Intellectual property rights, political risk and economic growth in developing countries. *Journal of Economics and International Finance* 1(6): 127-134.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth* 8: 155-194.

Barnett, S.M. & Williams, W. (2004). National intelligence and the emperor's new clothes. IQ and the wealth of nations. *Contemporary Psychology APA Review of Books* 49: 389-396.

Dickerson, R.E. (2006). Exponential correlation of IQ and the wealth of nations. *Intelligence* 34: 291-295.

Ginarte, J.C. & Park, W.G. (1997). Determinants of patent rights: A cross-national study. *Research Policy* 26: 283-301.

Hunt, E. (2011). *Human Intelligence*. Cambridge: Cambridge University Press.

Hunt, E. & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence* 36: 1-9.

Kanazawa, S. (2006). IQ and the wealth of states. *Intelligence* 34: 593-600.

Kanyama, I.K. (2014). Quality of institutions: Does intelligence matter? *Intelligence* 42: 44-52.

Kodila-Tedika, O. & Asongu, S.A. (2015). The effect of intelligence on financial development: A cross-country comparison. *Intelligence* 51: 1-9.

Lynn, R. & Vanhanen, T. (2012). *Intelligence. A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Meisenberg, G. (2012). National IQ and economic outcomes. *Personality and Individual Differences* 53: 103-107.

Meisenberg, G. (2014). Cognitive human capital and economic growth in the 21st century. In: T. Abrahams (ed.), *Economic Growth in the 21st Century: Perspectives, Role of Governmental Policies, Potential and Constraints*, pp. 49-106. New York: Nova Publishers.

Nechyba, T. (2004). Review of IQ and the Wealth of Nations. *Journal of Economic Literature* 42: 220-221.

Odilova, Sh. & Gu X. (2016). Patent protection, intelligence and economic growth: A cross-country empirical investigation. *Journal of Research in Business, Economics and Management* 6(1): 798-803.

Potrafke, N. (2012). Intelligence and corruption. *Economics Letters* 114: 109-112.

Ram, R. (2007). IQ and economic growth: Further augmentation of the Mankiw–Romer–Weil model. *Economics Letters* 94: 7-11.

Salahodjaev, R. (2015a). Democracy and economic growth: The role of intelligence in cross-country regressions. *Intelligence* 50: 228-234.

Salahodjaev, R. (2015b). Intelligence and finance. *Personality and Individual Differences* 86: 282-286.

Salahodjaev, R. (2015c). Intelligence and shadow economy: A cross-country empirical assessment. *Intelligence* 49: 129-133.

Salahodjaev, R. (2016). Intelligence and deforestation: International data. *Forest Policy and Economics* 63: 20-27.

Thompson, M.A. & Rushing, F.W. (1996). An empirical analysis of the impact of patent protection on economic growth. *Journal of Economic Development* 21: 61-79.

Volken, T. (2003). IQ and the wealth of nations: A critique of Richard Lynn and Tatu Vanhanen's recent book. *European Sociological Review* 19: 411-412.

Whetzel, D.L. & McDaniel, M.A. (2006). Prediction of national wealth. *Intelligence* 34: 449-458.

The Intelligence of Yakuts and Ethnic Russians in Yakutia

Vladimir Shibaev*
Institute of Psychology of Russian Academy of Sciences, Moscow, Russian Federation

Richard Lynn
University of Ulster, UK

*Corresponding author. E-mail: Wladimirdw@gmail.com

The intelligence of Yakuts and ethnic Russians in Yakutia was tested with Raven's Standard Progressive Matrices Plus. According to British norms, the average IQ was 97.0 for Yakuts and 97.9 for ethnic Russians. Urban samples scored higher than rural samples.
Key Words: SPM+, Intelligence, Yakuts, Russia.

The objective of this paper is to report a study of the intelligence of Yakuts and ethnic Russians in the province of Yakutia in eastern Siberia. Yakutia (also known as Sakha, the former indigenous term) is the largest of the 83 provinces of the Russian Federation. It consists of approximately 3,000,000 km^2 but has only a small population given in the 2010 census as approximately 958,500. The capital city Yakutsk has a population of approximately 269,600. The next largest cities are Neryungry with a population of approximately 58,000 and Mirny with a population of approximately 34,000. The population of the province is given in the 2010 census as 64 percent urban. Slightly above 50 percent of the economy is generated by industrial production consisting primarily of diamonds, gold and tin ore mining. The rest of the economy consists largely of agriculture. The GDP per capita of the province is 28th among the provinces of the Russian Federation. Education between the ages of 7 and 18 years is compulsory for all children in the Russian Federation. Yakutsk has a university – the North-Eastern Federal University – and several theatres. Many of the population have internet connection and TV.

The indigenous population are the Yakuts, also known as Turkic Sakha. They are classified by Cavalli-Sforza, Menozzi and Piazza (1996, p. 231) as one of the

Arctic peoples most closely related to the Sherpa and Tuva, and more distantly related to the peoples of the Altai Mountains, Northern Chinese and Uzbeks. They originated in Central Asia and migrated into Yakutia in the 13th and 14th centuries. Russia began to colonize Yakutia in the 17th century and a number of Russians subsequently migrated into the province. The 2010 census gives the population as Yakuts (49.9%), Russians (37.8%), Evenks (2.2%), Ukrainians (2.2%), Evens (1.6%) and Tartars (0.9%). Yakutia has a sub-arctic climate with average monthly temperatures in Yakutsk of minus 35ºC in January and 19ºC in July.

The study to be reported is the first of the intelligence of Yakuts and ethnic Russians in Yakutia, but there have been a few studies of the intelligence of ethnic Russians in European Russia and of other ethnic peoples in Eastern Siberia. Grigorenko and Sternberg (2001) have reported an IQ of 96 for a sample of Russian adults in European Russia tested with the Cattell Culture Fair test, and Lynn (2001) has reported an IQ of 97 for a sample of Russian 14-15 years olds in European Russia tested with the Standard Progressive Matrices. Lynn and Vanhanen (2012, p. 27) give an IQ of 96.7 for a sample of Russian 15-year-olds based on PISA results in math, science and reading comprehension, and an IQ of 96.6 for all three samples. All these are scaled to British norms, with a British mean and standard deviation of 100 and 15, respectively.

Studies of the intelligence of some of the ethnic minorities in Russia were carried out in the 1920s and 1930s by Luria and others. These have been reviewed by Grigoriev and Lynn (2009), who concluded that at that time these ethnic minorities had lower intelligence than ethnic Russians. These studies came to an end in 1936 when intelligence testing was banned in the Soviet Union. It was not until the 1960s and early 1970s that this prohibition was progressively relaxed and research on intelligence in Russia resumed. Shibaev and Lynn (2015) have reported a study of intelligence tested with the Standard Progressive Matrices Plus of a small sample of Evenk/Tungus school children in a village in an isolated rural location in the province of Vladivostok who obtained an IQ of 80.

Method

The samples were drawn from Yakutsk, the capital of the Yakutia region containing almost one third of the population of the province, from the town of Viluysk, and from four villages. The sample from Yakutsk consisted of 52 ethnic Russian and 287 ethnic Yakut school students with a mean age of 12 to 13 years attending typical secondary schools. They included all those present in the first year of secondary school on the day of testing.

Viluysk is situated 592 km north-west of Yakutsk and was officially designated as a city in 2013 but has a population of approximately 10,600 and would be regarded in western countries as a small town. There is no institution of tertiary education but there is a museum of local history and a museum of the Yakut's Jewish harp (the khomus). Most people have an internet connection. The sample from Viluysk consisted of 7 ethnic Russian and 236 ethnic Yakut school students with a mean age of 12.7 years attending the three secondary schools of the town. Two of these were typical secondary schools, and one was a gymnasium (academic-track secondary school). The sample included all those present in the first year of secondary school on the day of testing.

The village samples were:

1. *Khangalassy:* Located approximately 45 km north from Yakutsk, it has a population of approximately 1,800 ethnic Russians including a small number of "sakhalyars" of mixed Russian-Yakut ethnicity. It has one school. All children aged 7 to 16 were tested, with an average age of 10.8 (N=61).
2. *Khatassy:* This village is approximately 40 km from Yakutsk and consists of approximately 5,000 Yakut inhabitants. It has one school. All children aged 9 to 15 were tested, with an average age of 10.9 (N=75).
3. *Berdigestyakh:* This is the administrative center of Gorny Ulus (Gorny rural region) and is approximately 184 km from Yakutsk. The population is Yakut and numbers 6,460.There are two secondary schools in which all children aged 9 to 16 were tested, with an average age of 12.9 (N=82). There is also a gymnasium (academic-track secondary school) from which 21 children aged 15 to 17 with an average age of 16.75 were tested.
4. *Yert:* This village is approximately 264 km from Yakutsk and has a Yakut population of 544. There is only one school in this village. All children aged 7 to 18 were tested, with an average age of 12.6 (N=58).

The samples were tested between late 2015 and late 2016 with Raven's Standard Progressive Matrices Plus (Raven, 2008) administered by the first author with a time limit of one hour giving ample time to complete the test. Boys and girls attended the same schools and there were approximately equal numbers of each in all samples.

Results

The British-scaled IQs of the participants were obtained from the norms of the 2008 British standardization given in Raven (2008).The results for all secondary school students are shown in Table 1. The results for the village secondary school students are shown in Table 2. The results for gymnasium school students are shown in Table 3. The results for university students are shown in Table 4. It seems surprising that the IQs of pupils from gymnasiums are higher than those of students in Yakutsk University. One likely reason is that most of the university students were studying subjects such as philology and linguistics. Results from countries as diverse as Sudan and the United States have shown that students in the "soft sciences" and humanities usually have lower IQs than students in the natural sciences, mathematics and engineering (e.g., Khaleefa, Amer & Lynn, 2014). Another reason could be that pupils from gymnasiums are more motivated and disciplined than university students. Also, the best pupils from gymnasiums try to graduate in Moscow or St. Petersburg universities, not in the local Yakutsk University. The high-scoring pupils in Berdigestyakh's gymnasium were aged 10-11. It is not clear whether their very high scores are related to their young age or the process by which local children are selected into the gymnasium. It may also be a chance result due to the small size of the sample.

Table 1. *IQs of Russian and Yakut secondary school students.*

	Russians		Yakuts		
Sample	N	IQ ± SD	N	IQ ± SD	t
Yakutsk	52	102.3 ± 17.5	287	97.9 ± 15.8	1.69
Viluysk	7	97.1 ± 18.2	236	99.5 ± 13.7	0.35
Villages	61	91.2 ± 14.1	215	94.2 ± 15.6	1.43

Table 2. *IQs of Russian and Yakut village school students.*

Sample	Ethnicity	N	IQ ± SD
Khangalassy	Russians	61	91.2 ± 14.1
Khatassy	Yakuts	75	93.2 ± 16.1
Berdigestyakh	Yakuts	59	93.0 ± 17.4
Berdigestyakh	Yakuts	23	95.9 ± 12.9
Yert	Yakuts	58	96.1 ± 14.3

Table 3. *IQs of Yakut gymnasium students.*

Sample	Ethnicity	N	IQ ± SD
Viluysk	Yakuts	132	110.64 ± 13.82
Berdigestyakh	Yakuts	21	127.25 ± 13.61

Table 4. *IQs of Yakut first-year university students aged 17-20 years.*

Sample	Ethnicity	N	IQ ± SD
Yakutsk	Yakuts	46	101.16 ±15.34

Discussion

There are five points of interest in the results. First, the results given in Table 1 show that the British-scaled IQ of the urban Russians averaged from the two city samples weighted by sample size was 101.7 and of the village sample was 91.2. The population of the province is 64 percent urban and 36 percent rural. Thus, weighting the IQs of the urban and rural samples by their percentages in the province gives an IQ of 97.9. This is marginally higher than the IQ for Russia given as 96.6 by Lynn and Vanhanen (2012) on the basis of three studies from European Russia. These results are the first to show that the IQ of Russians is approximately the same in Asiatic Russia as in European Russia.

Second, the results given in Table 1 show that the IQ of the urban Yakuts averaged from the two samples weighted by sample size was 98.6 and of the village sample was 94.2. The population of the province is 64 percent urban and 36 percent rural. Thus, weighting the IQs of the urban and rural samples by their percentages in the province gives an IQ of 97.0.

Third, the Russians obtained a slightly higher IQ than the Yakuts in the Yakutsk sample while the Yakuts obtained slightly higher IQs than the Russians in the Viluysk and village samples, but none of these differences was statistically significant. The IQs of the combined samples are estimated as 97.0 (SD 15.1) for the Yakuts (N=738) and 97.9 (SD 16.5) for the ethnic Russians (N=120). This difference is trivial and not statistically significant (t=.56). Therefore the IQs of the two samples should be regarded as approximately the same. Weighting these IQs by the percentages of Yakuts and Russians in the province of 49.9 and 37.8, respectively, gives an IQ of 97.4 for the province.

Fourth, for both Russians and Yakuts the IQs of the city samples were higher than the IQs of the village samples. For the Russians, there was a difference of 10.5 IQ points between the combined city samples and the village sample, while for the Yakuts the difference was 4.4 points. The higher IQs of the city samples is a common result found in many previous studies reporting that urban populations typically obtain higher IQs than rural populations. In a study of intelligence in 79 provinces of the Russian Federation, a correlation of .44 with urbanization has been reported by Grigoriev et al. (2016). In the United States, in the standardization sample of the American WAIS-R, the IQ of the urban sample was 1.9 IQ points higher than that of the rural sample (Reynolds et al., 1987). The populations of capital cities typically obtain higher IQs than the rest of the population, e.g. in the British Isles (Lynn, 1979), France (Lynn, 1980) and Portugal (Almeida, Lemos & Lynn, 2011). The higher average IQs of urban samples are probably explained by migration of individuals with higher IQ to towns and cities for which evidence is given for the British Isles and France (Lynn, 1979, 1980) and Russia (Grigoriev et al., 2016). In Yakutia, there has been substantial rural-to-urban migration in the recent past, and also emigration from the province. Between the 1989 and 2010 censuses, the population of Yakutsk increased from 186,626 to 269,601 while the population of the province declined from about 1.1 million to 951,000. Because of out-migration of young working-age people, mainly old people are left in the villages. If those with higher IQ were more likely to migrate than those with lower IQ, this can explain the urban-rural IQ difference that we observed in our study.

Fifth, Yakutia has a sub-arctic climate with an average monthly temperature of minus 35°C in January and 19°C in July in Yakutsk. The high IQ of the Yakuts is therefore consistent with the 'cold winters theory' of the evolution of racial differences in intelligence advanced by Lynn (1991, 2015). The theory proposes that greater cognitive demands for survival in the temperate and cold environments of the northern hemisphere provided the crucial selection pressure for the evolution of higher intelligence. This theory has been endorsed by a number of scholars who have reported statistically significant associations between average winter temperatures and national IQ (Rushton, 2000; Kanazawa, 2008; Templer & Arikawa, 2006; Meisenberg & Woodley, 2013). The high IQ of the Yakuts is a further corroboration of this theory.

References

Almeida, L.S., Lemos, G.C. & Lynn, R. (2011). Regional differences in intelligence and per capita incomes in Portugal. *Mankind Quarterly* 52: 213-221.

Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1996). *The History and Geography of Human Genes*, abridged paperback edition. Princeton, NJ: Princeton University Press.

Grigorenko, E.L. & Sternberg, R.J. (2001). Analytical, creative and practical intelligence as predictors of self-reported adaptive functioning: A case study in Russia. *Intelligence* 29: 57-73.

Grigoriev, A. & Lynn, R. (2009). Studies of socioeconomic and ethnic differences in intelligence in the former Soviet Union in the early twentieth century. *Intelligence* 37: 447-452.

Grigoriev, A., Ushakov, D., Valueva, E., Zirenko, M. & Lynn, R. (2016). Differences in educational attainment, socio-economic variables and geographical location across 79 provinces of the Russian Federation. *Intelligence* 58: 14-17.

Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence* 36: 99-108.

Khaleefa, O., Amer, Z. & Lynn, R. (2014). IQ differences between arts and science students at the University of Khartoum. *Mankind Quarterly* 55: 136-146.

Lynn, R. (1979). The social ecology of intelligence in the British Isles. *British Journal of Social and Clinical Psychology* 18: 1-12.

Lynn, R. (1980). The social ecology of intelligence in France. *British Journal of Social and Clinical Psychology* 19: 325-331.

Lynn, R. (1991). The evolution of race differences in intelligence. *Mankind Quarterly* 32: 99-173.

Lynn, R. (2001). Intelligence in Russia. *Mankind Quarterly* 42: 151-154.

Lynn, R. (2015). *Race Differences in Intelligence: An Evolutionary Analysis*. Augusta, GA: Washington Summit Publishers.

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Meisenberg, G. & Woodley, M.A. (2013). Global behavioral variation: A test of differential-K. *Personality and Individual Differences* 55: 273-278.

Raven, J. (2008). *Standard Progressive Matrices-Plus Version and Mill Hill Vocabulary Scale Manual.* London: Pearson.

Reynolds, C.R., Chastain, R.L., Kaufman, A.S. & McClean, J.E. (1987). Demographic characteristics and IQ among adults. *Journal of School Psychology* 25: 323-342.

Rushton, J.P. (2000). *Race, Evolution and Behavior*. Port Huron, MI: Charles Darwin Research Institute.

Shibaev, V. & Lynn, R. (2015). The intelligence of the Evenk/Tungus of the Russian Far East. *Mankind Quarterly* 55: 202-207.

Templer, D.I. & Arikawa, H. (2006). Temperature, skin color, per capita income and IQ: An international perspective. *Intelligence* 34: 121-139.

MANKIND QUARTERLY 2017 57:4 687-705

Testing Spearman's Hypothesis with Alternative Intelligence Tests: A Meta-Analysis

Jan te Nijenhuis*, Michael van den Hoek, Denise Willigers
University of Amsterdam, Netherlands

*Corresponding author: Jan te Nijenhuis, Gouden Leeuw 746, 1103 KR, Amsterdam, the Netherlands: email: JanteNijenhuis@planet.nl

Spearman's hypothesis states that differences between groups on subtests of an IQ battery are a function of the *g* loadings of these subtests. We examined Spearman's hypothesis in alternative tests of intelligence. We compared various groups. We carried out a meta-analysis based on 20 data points and a total of 8,322 Whites and 2,507 minority group members. Spearman's hypothesis was strongly confirmed with a mean *r* of .62. We conclude that group differences are to the same degree a function of *g* when they are measured with traditional test batteries as when they are measured using alternative test batteries.
Key Words: Cross-cultural assessment, Spearman's hypothesis, *g*, Intelligence, Meta-analysis.

Differences between racial and ethnic groups in mean intelligence test scores are well-established findings (Rushton & Jensen, 2005). The magnitude of these differences, however, varies as a function of the type of cognitive skills the tests measure as tests that are stronger measures of general intelligence (*g*) exhibit larger mean differences than tests that are weaker measures of *g* (Jensen, 1998).

Charles Spearman (1927) initially observed that race differences should be "most marked in just those [tests] which are known to be saturated with *g*" (p. 379), an effect Jensen (1980) later named *Spearman's hypothesis*. Spearman and Jensen in his early papers focused on Black/White differences, but the term has expanded and now includes all kinds of group comparisons, for instance between Whites and Hispanics. There are three levels of Spearman's hypothesis that Jensen (2001) called the strong form, weak form, and the contra hypothesis.

The *strong form* posits that any observed race differences in test score means are only due to differences in *g*. The *weak form* posits that race differences in test score means are due to differences in *g* as well as differences in other aspects of cognitive ability (e.g., long-term retrieval, spatial ability). The *contra hypothesis* posits that race differences in test scores are only due to non-*g* aspects of cognitive ability or test specific factors (i.e., test bias).

Jensen (1985, 1998) carried out a series of classic studies on Spearman's hypothesis and found support for the weak form of the hypothesis. Te Nijenhuis and co-authors carried out meta-analyses on, respectively, Black adults/White adults comparisons (te Nijenhuis & van den Hoek, 2016), native Dutch/ethnic minority immigrants from non-Western countries comparisons (te Nijenhuis et al., 2016), and European Jews/non-Jewish Whites and European Jews/Oriental Jews comparisons (te Nijenhuis et al., 2014), and found support for the weak form of Spearman's hypothesis. Armstrong, Woodley and Lynn (2014) compared Lapps and Finns in Finland and found a zero correlation. Warne (2016) tested Spearman's hypothesis with advanced placement examination data and found support for its weak form. So, studies to date have generally found support for the weak form of the hypothesis: both *g* and other aspects of cognitive ability play a role in the observed test score differences between races (Jensen, 2001; Schönemann, 1997, but see Rushton, 2003). While differences in *g* have the largest influence, the other cognitive abilities that can influence mean test score differences are spatial ability (typically favoring White and Asian respondents) and memory (typically favoring Black and Hispanic respondents).

Assessing Spearman's Hypothesis

There are two methods used to assess Spearman's hypothesis: multi-group confirmatory factor analysis (MG-CFA), and correlated vectors (CV). The MG-CFA procedure assesses Spearman's hypothesis by conducting a confirmatory factor analysis (latent variable analysis) simultaneously on separate data from two or more comparison groups. If a certain level of invariance exists among the tests, then the differences found in the latent means and variances are used to assess Spearman's hypothesis (Dolan & Hamaker, 2001). While some argue that MG-CFA is a sound method for investigating Spearman's hypothesis, Woodley et al., (2014) argued it has two major drawbacks. First, MG-CFA requires large datasets, so studies that use small datasets cannot use this method to analyze their data. Second, because most of the original studies do not report sufficient information (e.g., means and covariances of both groups), MG-CFA cannot be used when conducting a meta-analysis. Consequently, they wrote that the CV approach is better for examining Spearman's hypothesis when meta-analytically combing data from multiple studies.

A correlated vectors analysis involves correlating the *g* loading of a group of tests with some other variable, such as the size of mean group differences in the tests (Jensen, 1998). A positive correlation indicates that tests with higher *g* loadings have larger group differences in mean test scores, although there is no agreed-upon correlation value that differentiates the strong and weak forms of Spearman's hypothesis.

Tests Used to Measure Spearman's Hypothesis

Studies examining Spearman's hypothesis typically use individually-administered multi-factor measures of cognitive ability, such as Wechsler and Kaufman intelligence scales. Some have argued that such tests are biased, thus any group differences are to some degree due to factors other than differences in cognitive ability (van de Vijver & Tanzer, 2004). Flanagan, Ortiz, and Alfonso (2013) go so far as to state that test score interpretations are invalid for "individuals whose experiences and development (not from intrinsic delays but rather circumstantial changes as might occur with respect to learning English as a second language) differ from those…established by the population on whom the test was normed" (p. 294). The two most common arguments for test bias come from differences in culture and language.

The argument that tests are culturally biased tends to use some variation of the following logic (Helms, 1992; Helms-Lorenz, van de Vijver & Poortinga, 2003). When test authors create a test of cognitive ability, they make cultural assumptions and references (whether implicit or explicit) in the test items and/or its administration. Examples of ways to possibly culturally bias tests are to: include items using pictorial objects as responses that are only familiar to respondents from certain cultures, require extensive language use, require specific knowledge that is more commonly available to one cultural group, or require the use of information learned outside of the test (Jensen, 1980). The amount of culture embedded in a test is often referred to as the test's "cultural loading."

The cultural test bias hypothesis has been examined for over 30 years and has largely been found to not be a major cause of group differences (Brown, Reynolds & Whitaker, 1999). Nonetheless, scholars still argue that this is a major cause of between-group differences in test scores (for a review, see Frisby, 2013). As their arguments tend to come from a philosophical stance rather than an empirical one, empirical studies showing lack of bias are likely not going to appease them (e.g., Helms, 2006).

The issue of language bias comes from the idea that examinees whose first language is different from the language used in the test perform worse because of differences in language fluency. Thus, even though the test was designed to measure cognitive ability, it is also measuring a second construct (Ackerman, 1992)—what Cummins (1979) called "academic language proficiency" (i.e., language skills needed to be successful in academic settings). Consequently, differences in language proficiency, in addition to—or in lieu of—differences in cognitive ability, are causing group differences in mean test scores. This can occur even if using a translator or the original test was translated into the examinee's first language (Hambleton & Li, 2005). There is clear empirical proof that language bias plays a role, for instance when testing first-generation non-Western immigrants (te Nijenhuis & van der Flier, 2003; te Nijenhuis et al., 2016).

Alternative Measures of Cognitive Ability

The criticisms of traditional tests of intelligence have been around for over a century and were the impetus for developing alternative, culture/language-reduced measures of cognitive ability (e.g., Cattell, 1940; Yerkes, 1921). Such measures were created because of the limitations associated with assessing a diverse population (McCallum, 2013). Culture-reduced measures of intelligence are typically designed to measure *g*, but do so without relying as heavily on language or culture as traditional intelligence tests (Braden & Athanasiou, 2005).

To date, studies of Spearman's hypothesis have typically not used alternative measures of cognitive ability. Such a study could be useful in understanding differences in cognitive ability. Do group differences in intelligence show the same patterns for alternative test batteries as they do for traditional test batteries? A recent meta-analysis shows that traditional test batteries with language-biased subtests show no strong confirmation of Spearman's hypothesis and that after leaving these biased subtests out the correlation becomes substantially higher (te Nijenhuis et al., 2016). Another recent meta-analysis on differences between Black and White adults shows a meta-analytical $r = .57$. Quite a few alternative test batteries contain subtests with a substantial language component, and we expect this to lead to a low correlation between *g* loadings and group differences. We also expect that when these subtests are left out the correlation becomes substantially larger. We make an exception for White-Northeast Asian comparisons, where we expect there will either be a negligible or negative correlation between group differences and the subtests' *g* loadings, as this is found in the literature (Nagoshi et al., 1984)

Method

To test our hypotheses, we used a meta-analysis. In a meta-analysis, studies that address similar questions are quantitatively combined (Schmidt & Hunter, 2015). The typical steps when conducting a meta-analysis are: (a) search the literature; (b) decide on criteria for including a study; (c) calculate the effect sizes; (d) combine the studies' effect sizes; (e) rule out rival explanations; and (f) report the findings (Durlak, 1999; Field and Gillett, 2010). Meta-analyses are very important in understanding psychological phenomena. Not only can they aid in understanding the generalizability or reproducibility of a finding, but they can also overcome some of the biases associated with studies that use small samples. We decided to use Pearson *r* to compute the correlation between group differences (*d*) and *g* loadings, following earlier-conducted meta-analyses using Pearson *r* in the method of correlated vectors (e.g. te Nijenhuis, van Vianen & van der Flier, 2007).

Literature Search

Several search strategies were used. For the digital search the following electronic databases were used: Google Scholar, ProQuest, PsycINFO, and CataloguePlus (Primo). The terms used to find studies were 'culture free', 'culture reduced', 'culture fair' and 'non-verbal' in conjunction with the terms: 'intelligence', 'IQ', 'mental ability', 'mental capacity', 'cognitive ability', 'aptitude', 'competence', and 'differences'. References in important papers were also followed up for more studies and data.

Inclusion Criteria

We used three inclusion criteria for this meta-analysis. First, the study had to use at least four cognitive ability tests/subtests. Second, the study's participants were not selected on a highly *g*-loaded variable (e.g., gifted, referred for special education). Third, the study used alternative intelligence tests and the study authors stressed that the test was culture- or language-reduced.

Effect Sizes: Mean Differences

The effect size typically used to measure between-group differences in Spearman's hypothesis studies is *d*, which is an estimate of the standardized difference between groups. For the current study we followed the procedures established by Jensen, so *d* was calculated by subtracting the score of the lower scoring group from the score of the higher scoring group and then dividing the difference by a standard deviation. So, for instance, the scores of the White group were subtracted from the scores of the Northeast Asian group and not the other way round.

The selection of standard deviations is very important for calculating the correct effect size, the standardized difference between groups. Therefore a procedure was used to select which standard deviation would be best used in these calculations. Whenever possible, the standard deviations were taken from nationally representative standardization samples or norming samples for the tests used in the study. This is the preferred option because the standard deviation of a large and representative sample is closer to the population standard deviation than a small study sample. For some tests, however, no such samples were available. When this was the case, the standard deviations of the largest group were used to compute *d*, since a larger group would still have more reliable standard deviations than a small group. If both groups were of equal size, the *SD* from the majority group was used because it is more likely to be representative of the population *SD* than that of the minority group.

g *Loadings*

Our first choice when calculating tests' *g* loadings was to use data from standardization samples, but some studies used a battery of tests for which standardization data were not available. In such situations, we tried to find a study that matched the sample from the meta-analytical data point using a large and representative sample. If no such studies existed, we used datasets from studies with smaller sample sizes, or used the correlations or *g* loadings reported for the meta-analytical data point.

We estimated the *g* loadings in a study using the first unrotated factor from a principal axis factor analysis. While this is not always the optimal way of extracting *g*, it often produces *g*-loadings similar to more refined methods (Jensen & Weng, 1994). Moreover, it can be used when the only data available are test correlations, which is often the case when using summary statistics from a published study.

Average Correlation across Samples

To compute the average correlation, we followed the procedures outlined by Schmidt and Hunter (2015), which uses the untransformed correlations to calculate the weighted mean effect size. Each correlation was weighted using the harmonic *N* (see below for more details). We calculated the weighted mean correlation using a random effects model. We chose a random effects model because we believe the population of Spearman's hypothesis studies share a common mean effect, but the studies for our particular analysis are not exact replicates of each other (Hedges & Vevea, 1998). Moreover, it produces accurate aggregate estimates, except when the population value is extremely large (Field, 2001).

Sample Weights

When combining studies in a meta-analysis, it is important to weight the studies so that they make an appropriate contribution. A study with 1000 participants should make a larger contribution to the final effect size than a study with 100 participants. The current meta-analysis involved combining studies comparing already-formed groups and there was often a large disparity between group sizes. For example, Sternberg et al.'s (2006) sample contained 348 White and 47 Black participants. The sample of 47 has quite substantial sampling error, whereas a sample of 348 produces a much smaller sampling error. Combining the sample sizes without correction could produce a spurious estimate of the study's precision.

An elegant way to combine the groups' sample sizes when they are so disparate is to calculate a harmonic *N* (Klockars & Sax, 1987). While there are different formulae for a harmonic *N*, we use the one from te Nijenhuis and van der Flier (2003), which is shown in Equation 1.

$$\frac{G \times G}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots \frac{1}{x_n}}$$

where *G* is the number of groups and x_i is the size of each individual group. The advantage of this formula is that it produces a conservative *N* for a study with disparate samples sizes, but produces an uncorrected *N* for a study with equal sample sizes. For example, sample sizes of 348 and 47 produce a harmonic *N* of 166, while a study with samples of 100 and 100 produces a harmonic *N* of 200.

Moderators

In a meta-analysis, it is important to determine whether the collected effect sizes comprise a homogeneous or heterogeneous group (Durlak, 1999). If the effect sizes are heterogeneous, then it is important to determine if the between-study variability in effect sizes is due to sampling error or due to some substantive variable. To examine study heterogeneity, we calculated the amount of variance explained by sampling error.

We expected that there might be two causes of heterogeneity in our analysis. First, as we indicated in our hypotheses, we expect that all the correlations will be positive except when the comparison group is comprised of Asians. When using Asians in the comparison group, we expect the correlation to be negative. A second possible cause of heterogeneity is language bias against a group of interest (Jensen, 1998).

Determining Language Bias

Jensen (1998, p. 372-374) describes some requirements for testing Spearman's hypothesis, and the fourth is that any test that is demonstrably biased (in the psychometric/statistical sense) with respect to the groups of interest should be excluded. This is clearly meant to keep subtests with language bias out. We expected to only find language-biased subtests with immigrant populations. Te Nijenhuis and van der Flier (2003) and te Nijenhuis et al. (2016) already reported on the topic of correcting for language bias using Spearman's hypothesis, but we choose a slightly different approach using an explicit cutoff score. First, attention was focused on subtests with a substantial language component. A first scatter plot with *g* loadings on the x-axis and *d* scores on the y-axis was drawn and it was checked whether the subtests with a substantial language component were substantially above the regression line. These subtests were then left out for the second scatter plot, and a new regression line of *d* on *g* was drawn. The expected value of *d* for the possibly biased subtest was recalculated using the *g* loading of the subtest. The expected *d* score was then subtracted from the observed *d* score, and if the residual value was .20 or larger, the subtest was considered to be language biased. This differs slightly from the approach taken by te Nijenhuis et al. who did not use an explicit cutoff score. The new approach has the advantage of offering a more explicit criterion.

Software

We corrected for sampling error using the meta-analytical software developed by Schmidt and Le (2004). The program enables a computation of the amount of variance between the studies that can be explained by sampling error.

Samples

The literature search yielded six studies that met the inclusion criteria and 20 different group comparisons. Across the six studies, there were 8,322 Whites and 2,507 minority group members, yielding a total harmonic *N* of 6,661. The studies are listed in Table 1 and are described in the following sections.

1. Sternberg et al. (2006)

These authors studied indicators of creative and practical skills to supplement the SAT in predicting college GPA as well as reduce between-race score differences. The sample in the study consisted of 47 Blacks, 89 Hispanics, 348 Whites, 77 Asians, 11 Pacific Islanders, 11 Native Americans, 37 describing themselves as "other", and 157 did not report an ethnic background. The samples we used in this study are the Asian, Black, Hispanic, Native American and White samples.

The supplemental measures used in this study were based on Sternberg's theory of successful intelligence, which posits that skills other than analytical intelligence are important for success in life (Sternberg, 2002). The supplemental tests came from Sternberg's Triarchic Abilities Tests (STAT) as well as additional tests of creative and practical skills. As these supplemental tests were designed to assess different aspects of intelligence, we treated them all as alternative tests of cognitive ability.

The STAT is a nine-subtest battery designed to measure analytical, practical, and creative skills using multiple-choice items (Sternberg et al., 1996). Each subscale includes 5 items for a total of 45 items. These items are combined to produce three aggregate scores: Analytical, Practical, and Creative. The additional measures of creative skills were: (a) *Cartoons*: respondents select three of five *New Yorker* cartoons without captions and fill in the captions, which are then rated for creativity on a 5-point scale by trained judges; (b) *Written Stories*: participants select two of six story titles and then write a short story appropriate for each of the two titles in 15 minutes, which are rated for originality, complexity, emotional evocativeness, and descriptiveness on 5-point scales by trained judges; and (c) *Oral Stories*: respondents select two of five sheets of paper, each with 11-13 images that are linked by a common theme, after which they are given 15 minutes to formulate a short story for each paper and dictate it using a recording device, which are then rated for originality, complexity, emotional evocativeness, and descriptiveness on 5-point scales by trained judges. The additional measures of practical skills were: (a) *Everyday Situational Judgment Inventory*: respondents watch seven brief video vignettes with problems encountered in everyday situations and then select one of six written options as the most appropriate response to the situation; (b) *Common Sense Questionnaire*: respondents read 15 vignettes of business-related problems and select the most appropriate response to each vignette from 8 written options; and (c) *College Life Questionnaire*: respondents read 15 vignettes designed to capture common college-related situations and select the most appropriate response from several written options (the number of options varies depending on the situation).

2. Kane (2007)

Kane (2007) describes a study in which the Universal Nonverbal Intelligence Test (UNIT; Bracken, & McCallum, 1998) was administered to 77 White, 77 Black, 77 Asian, and 77 Hispanic participants. The UNIT is a norm-referenced test of intelligence that requires no verbal communication from the examiner or examinee. The test was standardized on a stratified sample of 2,100 children between the ages of 5 and 17 years that were designed to be representative of the United States population. The UNIT consists of 6 subtests measuring memory and reasoning using symbolic items that require internal verbal mediation using whatever idiosyncratic language system the examinee possesses, and non-symbolic items.

3. van den Berg (2001)

van den Berg (2001) administered the Multicultural Capacities Test of Middle Level [Multiculturele Capaciteiten Test Middelbaar niveau] (MCT-M; Bleichrodt & van den Berg, 2004) to natives and immigrants in the Netherlands. There were 857 native White participants as well as four groups of first-generation immigrants: 537 Surinamese, 150 Antillean, 342 Moroccan, and 191 Turkish. In addition, there was a group of 135 second-generation immigrants from a variety of ethnic backgrounds.

The MCT-M is a Dutch intelligence test designed to reduce bias in test scores of individuals coming from diverse cultural and ethnic backgrounds. The test was standardized on a sample of 9,352 adults with different educational backgrounds, of which 60.1% were immigrants. The test consists of eight subtests measuring comprehension knowledge, quantitative skills, fluid reasoning, spatial ability, and processing speed.

4. Naglieri, Rojahn & Matto (2007)

These authors administered the Cognitive Assessment System (CAS; Naglieri & Das, 1997) to 244 Hispanic and 1956 non-Hispanic children and adolescents. The non-Hispanic group was comprised of White (80.8%), Black (15.2%), Asian (3.4%), Native American (0.3%), and other (0.4%) races.

The CAS is designed to measure different aspects of cognitive processing based on the Planning, Attention, Simultaneous, and Successive (PASS) intelligence theory (Naglieri & Das, 2005). It was standardized on 2,200 children and adolescents between the ages of 5 and 17. There are 12 subtests, which are designed to measure each of the 4 PASS domains: Planning, Simultaneous, Attention, and Successive processes.

5. Hessels (1993)

Hessels (1993) administered the Learning Potential Test for Ethnic Minorities (LEM; Hessels & Hamers, 1993) to 400 children whose parents were both born in Turkey or Morocco as well as a group of 100 native Dutch children, divided in two age groups.The LEM was designed to measure cognitive ability of Turkish and Moroccan children in the Netherlands between the ages of five and eight years. Although based on the theory of learning potential, the LEM does not have a pretest or explicit training phase that other learning potential measures include. Instead, training is incorporated into each subtest so there is only one score per test. The LEM contains six different subtests measuring inductive reasoning, paired-associate learning, and short-term memory.

6. Flemmer & Roid (1997)

Flemmer and Roid (1997) report the scores of 258 White and 62 Hispanic adolescents on the Reasoning and Visualization subtests of the Leiter International Performance Scale-Revised (Leiter-R; Roid & Miller, 1997). According to Flemmer and Roid (1997), the main advantage of the Leiter-R is that the subtests are entirely non-verbal, both in responding and giving directions to the participants. The subjects are only required to give one of three possible responses, namely pointing, aligning objects, or aligning cards with pictures on them. Furthermore, the directions themselves are also given non-verbally, using gestures and signals. The Reasoning subscale contains a total of four subtests: Classification, Sequential Order, Repeated Patterns, and Design Analogies. The six Visualization subtests are: Matching, Figure-Ground, Paper Folding, Figure Rotation, Picture Context, and Form Completion. The Matching, Classification, and Picture Context subtests were not used because *g* loadings of sufficient quality could not be found in the literature.

Results

Meta-Analytic Results

The correlations between the *g* loadings and the score differences (*d*) are shown in Table 1. Most of the correlations are positive in sign and the large majority of them are substantial in magnitude. Table 2 presents the results of the meta-analysis of the 20 comparisons, and includes the number of correlation coefficients (*K*), total sample size (*N*), the mean-weighted vector correlation (mean *r*), and the standard deviation of the vector correlation (SD_r). The last column presents the percentage of variance explained by sampling error (%VE). The analysis of all 20 data points yields a mean vector correlation of .26, with a negligible 1.6% of the variance in the observed correlations explained by sampling error. Consequently, we next examined if there was an effect of outliers – data points that are distant from the distribution of the large majority of other data points – and an effect of language bias.

Table 1. *Summary of studies used in the meta-analysis.*

Study	Age (Range)	Test[c]	$N_{subtests}$	Comparison Groups[a]	$N_{majority}$	$N_{minority}$	$N_{harmonic}$	r (d x g)	r (d x g)$_{corrected}$
Hessels (1993)	6 (5.4-6.9)	LEM	6	Native White–Turkish	60	120	160	.87	.87
		LEM	6	Native White–Moroccan	60	120	160	.80	.96
	7.3 (6.10-7.9)	LEM	6	Native White–Turkish	40	80	107	-.49	-
		LEM	6	Native White–Moroccan	40	80	107	.25	.80
van den Berg (2001)	29 (16-60)	MCT-M	6	Native White–Surinam Immigrants	857	537	1321	.56	.68
		MCT-M	6	Native White–Netherlands Antilles Immigrants	857	150	511	.46	.91
		MCT-M	6	Native White–Moroccan Immigrants	857	342	978	.35	.52
		MCT-M	6	Native White–Turkish Immigrants	857	191	625	.29	.61
		MCT-M	6	Native White–Second Generation Immigrants	857	135	467	.32	.32
Naglieri et al. (2007)	8.3 (5-17)	CAS	12	Non-Hispanic–Hispanic	1956	244	868	-.47	.54
Sternberg et al. (2006)	20 (18-22)[b]	Various	11	White–Black	348	47	166	.46	.46
		Various	11	White–Hispanic	348	89	283	.60	.85
		Various	11	White–Native American	348	11	43	.76	.79
		Various	11	Asian–White	348	77	252	-.60	-.60
Kane (2007)	10.5	UNIT	6	White–Black	77	77	154	.53	.53
		UNIT	6	White–Hispanic	77	77	154	.41	.41
		UNIT	6	Asian–White	77	77	154	-.48	-.48
Flemmer & Roid (1997)	11-21	Leiter-R	7	Low Education White–Hispanic	17	17	34	.81	.81
			7	Medium Education White–Hispanic	72	24	72	.16	.16
			7	High Education White–Hispanic	169	12	45	.16	.16

Note. a: Higher scoring group listed first. b: Estimated age. c: See text for description of tests. $N_{harmonic}$ computed using the formula in Equation 1. r (d x g): correlation between g

loadings and standardized group differences. $r\ (d \times g)_{corrected}$: correlation between *g* loadings and standardized group differences corrected for language bias.

Asians as an Outlier?

A cursory glance at Table 1 leads to the conclusion that the two data points of the Asian samples are outliers. Table 2 shows that leaving out these data points yields a somewhat higher correlation of .31, with 1.8% of the variance in the observed correlations explained by sampling error.

Assessment for Language Bias

In several datasets we found language bias for the minority groups and adjusted accordingly. From the data of Hessels (1993) we removed Repeating Syllables from the LEM as a language-biased subtest for the Moroccan samples. From the data of van den Berg (2001) we removed Word Relations and Word Analogies as language-biased subtests for the Antillean, Moroccan, and Turkish sample. From the data of Naglieri, Rojahn and Matto (2007) we removed Verbal Spatial Relations, Word Series, Sentence Repetition, and Sentence Questions from the CAS as language-biased subtests. From the data of Sternberg et al. (2006) we removed the SAT-V, SAT-M, Cartoons, and Oral Stories subtests for Hispanics and Cartoons and STAT - Creative subtests for the Amerindians as language-biased subtests. Table 2 shows that adjusting for language bias yields a much higher correlation of .62, with still only 3.4% of the variance in the observed correlations explained by sampling error.

Table 2. *Summary statistics for total sample in meta-analysis.*

Studies Included	*K*	*N*	Mean *r*	*SDr*	%VE
All studies	20	6,661	.26	.41	1.6
All studies minus 2 outliers	18	6,225	.31	.13	1.8
All studies minus 2 outliers and minus language bias	17	6,148	.62	.18	3.4

Note. *K* = number of correlations; *N* = total sample size; Mean *r* = mean-weighted vector correlation; SD_r = standard deviation of observed correlation; %VE = percentage of variance accounted for by sampling errors.

Discussion

In this study, we examined Spearman's hypothesis in alternative tests of cognitive ability. Such measures are typically designed to measure *g*, but are described as doing so without relying as heavily on language or culture as traditional intelligence tests. As we hypothesized, there was a positive correlation between the Black/Hispanic/Native American-White differences and the subtests' *g* loadings, as well as a positive correlation between the immigrant-native group differences and the subtests' *g* loadings. However, the size of the meta-analytical correlation was only modest. As we hypothesized, there was a negative correlation between the Asian-White differences and the subtests' *g* loadings. The two data points on Asians both showed strong negative correlations. After excluding the quite substantial number of language-biased subtests, we found an average sample-size-weighted correlation of .62. So, there does seem to be virtually no difference in whether the test is traditional or alternative as the average Spearman's hypothesis correlation reported by Jensen (1998, p. 378) was .63. This means that group differences are to the same degree a function of *g* when they are measured with traditional test batteries as when they are measured using alternative test batteries.

A conclusion that alternative tests are not needed does not follow from the results. In fact, it can be persuasively argued that alternative tests still do a good job because they measure the same constructs as traditional tests. Again, due to the indifference of the indicator, each subtest/battery measures Spearman's *g*. This is true even for the CAS, which openly disavows Spearman's *g*. Because two very different tests measure the same underlying cardinal construct, it doesn't necessarily follow that one test must be redundant or expendable. A simple example clarifies this point. The Vocabulary and Matrix Reasoning subtests of the Wechsler scales demonstrate strong *g* loadings, on the order of .75. However, it would be premature and inaccurate to conclude that the subtests are identical or interchangeable, simply because they have similar g loadings. Obviously, Matrix Reasoning would provide a better measure of *g* if the examinee did not comprehend the language demanded by the Vocabulary subtest. If an argument must be made that some alternative tests are not the first choice for an assessment, an important one is that some alternative test batteries are based on theory that is less well supported by independent empirical study. For example, neither Sternberg's Triarchic Theory or the PASS system appears to hold up under empirical scrutiny, whereas classical test batteries are based upon well-supported hierarchical intelligence models (Carroll, 1993) or can be described well using a hierarchical intelligence model. Another argument that could be made against alternative test batteries is that some of them contain language-biased subtests, so in that respect they are not superior to traditional IQ batteries.

Acknowledgement

Thanks to Henk van der Flier for statistical advice. Thanks to Alex Beaujean for his excellent feedback on a previous version of this manuscript. Thanks to an anonymous reviewer for the constructive feedback which much improved the paper.

References

References marked with an asterisk were included in the meta-analysis.

Ackerman, T.A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement* 29: 67-91. doi: 10.2307/1434777

Armstrong, E.L., Woodley, M.A. & Lynn, R. (2014). Cognitive abilities amongst the Sami population. *Intelligence* 46: 35-39.

Bleichrodt, N. & van den Berg, R.H. (2004). *Handleiding Multiculturele Capaciteiten Test Middelbaar Niveau* [Manual Multicultural Capacity Test Middle Level]. Amsterdam: NOA.

Bracken, B.A. & McCallum, R.S. (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.

Braden, J.P. & Athanasiou, M.S. (2005). A comparative review of nonverbal measures of intelligence. In: D.P. Flanagan & P.L. Harrison (eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 2nd ed., pp. 557-577. New York, NY: Guilford.

Brown, R.T., Reynolds, C.R. & Whitaker, J.S. (1999). Bias in mental testing since *Bias in Mental Testing*. *School Psychology Quarterly* 14: 208-238. doi: 10.1037/h0089007

Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.

Cattell, R.B. (1940). A culture-free intelligence test. I. *Journal of Educational Psychology* 31: 161-179. doi: 10.1037/h0059043

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism* 19: 197-202.

Dolan, C.V. & Hamaker, E.L. (2001). Investigating black-white differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC, and a critique of the method of correlated vectors. In: F. Columbus (ed.), *Advances in Psychological Research*, Vol. 6, pp. 31-60. Huntington, NY: Nova Science Publishers.

Durlak, J.A. (1999). Meta-analytic research methods. In: P.C. Kendall, J.N. Butcher & G.N. Holmbeck (eds.), *Handbook of Research Methods in Clinical Psychology*, 2nd ed., pp. 419-429. Hoboken, NJ: Wiley.

Field, A.P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods* 6: 161-180. doi: 10.1037/1082-989X.6.2.161

Field, A.P. & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology* 63: 665-694. doi: 10.1348/000711010X502733

Flanagan, D.P., Ortiz, S.O. & Alfonso, V.C. (2013). *Essentials of Cross-Battery Assessment*, 3rd edition. Hoboken, NJ: Wiley.

*Flemmer, D.D. & Roid, G.H. (1997). Nonverbal intellectual assessment of Hispanic and speech-impaired adolescents. *Psychological Reports* 80: 1115-1122.

Frisby, C.L. (2013). Testing, assessment, and cultural variation: Challenges in evaluating knowledge claims. In: D.H. Saklofske, C.R. Reynolds & V.L. Schwean (eds.), *The Oxford Handbook of Child Psychological Assessment*, pp. 150-169. New York, NY: Oxford University Press.

Hambleton, R.K. & Li, S. (2005). Translation and adaptation issues and methods for educational and psychological tests. In: C.L. Frisby & C.R. Reynolds (eds.), *Comprehensive Handbook of Multicultural School Psychology*, pp. 881-903. New York, NY: John Wiley.

Hedges, L.V. & Vevea, J.L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods* 3: 486-504. doi: 10.1037/1082-989X.3.4.486

Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist* 47: 1083-1101. doi: 10.1037/0003-066X.47.9.1083

Helms, J.E. (2006). Fairness is not validity or cultural bias in racial-group assessment: A quantitative perspective. *American Psychologist* 61: 845-859. doi: 10.1037/0003-066X.61.8.845

Helms-Lorenz, M., van de Vijver, F.J.R. & Poortinga, Y.H. (2003). Cross-cultural differences in cognitive performance and Spearman's hypothesis: *g* or *c*? *Intelligence* 31: 9-29. doi: 10.1016/S0160-2896(02)00111-3

*Hessels, M.G.P. (1993). *Leertest voor etnische minderheden: Theoretische en empirische verantwoording*. [Learning Test for Ethnic Minorities: Theoretical and empirical justification.] Doctoral dissertation, University of Utrecht. Utrecht: The Netherlands.

Hessels, M.G.P. & Hamers, J.H.M. (1993). A learning potential test for ethnic minorities. In: J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijssenaars (eds.), *Learning Potential Assessment: Theoretical, Methodological and Practical Issues*, pp. 285-312. Lisse, Netherlands: Swets.

Jensen, A.R. (1980). *Bias in Mental Testing*. New York, NY: Free Press.

Jensen, A.R. (1985). The nature of the Black–White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences* 8: 193-219.

Jensen, A.R. (1998). *The g Factor: The Science of Mental Ability.* Westport, CT: Praeger.

Jensen, A.R. (2001). Spearman's hypothesis. In: J.M. Collis & S. Messick (eds.), *Intelligence and Personality: Bridging the Gap in Theory and Measurement*, pp. 3-24. Mahwah, NJ: Erlbaum.

Jensen, A.R. & Weng, L.-J. (1994). What is a good *g*? *Intelligence* 18: 231-258. doi: 10.1016/0160-2896(94)90029-9

*Kane, H. (2007). Group differences in nonverbal intelligence: Support for the influence of Spearman's *g*. *Mankind Quarterly* 48: 65-82.

Klockars, A.J. & Sax, G. (1987). *Multiple Comparisons.* Newbury Park, CA: Sage.

McCallum, R.S. (2013). Assessing intelligence nonverbally. In: K.F. Geisinger, B.A. Bracken, J.F. Carlson, J.I.C. Hansen, N.R. Kuncel, S.P. Reise & M.C. Rodriguez (eds.), *APA Handbook of Testing and Assessment in Psychology, Vol. 3: Testing and Assessment in School Psychology and Education*, pp. 71-99. Washington, DC: American Psychological Association.

Naglieri, J.A. & Das, J.P. (1997). *Cognitive Assessment System*. Itasca, IL: Riverside.

Naglieri, J.A. & Das, J.P. (2005). The PASS theory. In: D.P. Flanagan & P.L. Harrison (eds.), *Contemporary Intellectual Assessment*, 2nd ed., pp. 120-135. New York, NY: Guilford.

*Naglieri, J.A., Rojahn, J. & Matto, H.C. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence* 35: 568-579. doi: 10.1016/j.intell.2006.11.001

Nagoshi, C.T., Johnson, R.C., DeFries, J.C., Wilson, J.R. & Vandenberg, S.G. (1984). Group differences and first principal-component loadings in the Hawaii Family Study of

Cognition: A test of the generality of 'Spearman's Hypothesis'. *Personality and Individual Differences* 5: 751-753.

Roid, G.H. & Miller, L. (1997). *Leiter International Test of Intelligence—Revised*. Chicago: Stoelting.

Rushton, J.P. (2003). Race differences in *g* and the "Jensen Effect." In: H. Nyborg (ed.), *The Scientific Study of General Intelligence: A Tribute to Arthur R. Jensen*, pp. 147-186. Amsterdam, Netherlands: Pergamon.

Rushton, J.P. & Jensen, A.R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law* 11: 235-294. doi: 10.1037/1076-8971.11.2.235

Schmidt, F.L. & Hunter, J.E. (2015). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 3rd edition. Thousand Oaks, CA: Sage.

Schmidt, F.L. & Le, H. (2004). Software for the Hunter-Schmidt meta-analysis methods. University of Iowa, Department of Management & Organization, Iowa City, IA, 42242.

Schönemann, P.H. (1997). Famous artefacts: Spearman's hypothesis. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition* 16: 665–694.

Spearman, C.E. (1927). *The Abilities of Man: Their Nature and Measurement*. New York, NY: Blackburn Press.

Sternberg, R.J. (2002). Beyond *g*: The theory of successful intelligence. In: R.J. Sternberg & E.L. Grigorenko (eds.): *The General Factor of Intelligence: How General Is It?*, pp. 447-479. Mahwah, NJ: Erlbaum.

*Sternberg, R.J. & the Rainbow Project Collaborators (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence* 34: 321-350. doi: 10.1016/j.intell.2006.01.002

Sternberg, R.J., Ferrari, M., Clinkenbeard, P. & Grigorenko, E.L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly* 40: 129-137. doi:10.1177/001698629604000303

te Nijenhuis, J., David, H., Metzen, D. & Armstrong, E.L. (2014). Spearman's hypothesis tested on European Jews vs non-Jewish Whites and vs Oriental Jews: Two meta-analyses. *Intelligence* 44: 15-18.

te Nijenhuis, J. & van den Hoek, M. (2016). Spearman's hypothesis tested on Black adults: A meta-analysis. *Journal of Intelligence* 4: 6. doi:10.3390/jintelligence4020006 te Nijenhuis, J. & van der Flier, H. (2003). Immigrant-Majority group differences in cognitive performance: Jensen effects, cultural effects, or both? *Intelligence* 31: 443-459.

te Nijenhuis, J., van Vianen, A.E.M. & van der Flier, H. (2007). Score gains on *g*-loaded tests: No *g. Intelligence* 35: 283–300.

te Nijenhuis, J., Willigers, D., Dragt, J. & van der Flier, H. (2016). The effects of language bias and cultural bias estimated using the method of correlated vectors on a large database of IQ comparisons between native Dutch and ethnic minority immigrants from non-Western countries. *Intelligence* 54: 117-135.

van de Vijver, F. & Tanzer, N.K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* 54: 119-135. doi: 10.1016/j.erap.2003.12.004

*van den Berg, R.H. (2001). *Psychologisch onderzoek in een multiculturele samenleving* [Psychological research in a multicultural society]. Academic thesis. Amsterdam: Nederlands Onderzoekscentrum Arbeidsmarkt & Allochtonen.

Warne, R.T. (2016). Testing Spearman's hypothesis with advanced placement examination data. *Intelligence* 57: 87-95.

Woodley, M.A., te Nijenhuis, J., Must, O. & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from *g*: The return of the Brand effect. *Intelligence* 43: 27-34. doi: 10.1016/j.intell.2013.12.004

Yerkes, R.M. (1921). *Memoirs of the National Academy of Sciences: Psychological Examining in the United States Army.* (Vol. 15). Washington, DC: Government Printing Office.

About *Mankind Quarterly*

Mankind Quarterly was founded as a quarterly journal of anthropology, in the broadest sense of "the science of man," in 1961. This was a time when the study of man had already diversified into physical anthropology, ethnography, quantitative cross-cultural research, archaeology and other subspecialties.

These developments took place against the background of a widening gulf between the biological and social sciences. Following the leading dogma of the day, cultural and social anthropologists in academe had begun to deny the importance of biology for behavioral and cultural phenomena. Conversely, biological (physical) anthropologists aligned themselves with the "hard" sciences, many describing themselves as human biologists rather than anthropologists in an attempt to distance themselves from a social anthropology that they no longer saw as scientifically sound. In many places, these divisions persist to the present day. *Mankind Quarterly* was founded as a response to these centrifugal trends. Its founders believed in the interdependence of human biology, behavior and culture, and they understood biological and cultural diversity as the outcomes of evolutionary, ecological, and historic processes.

In short, *Mankind Quarterly* was established as a journal for those scholars who still believed in a unified "science of man" that studies the interactions between biological and cultural diversity. It was first published in Edinburgh, Scotland in 1961, and then, from 1979 to 2014, by the Council for Social and Economic Studies (chaired by Roger Pearson) in Washington, D.C. In January 2015, publication was transferred to the Ulster Institute for Social Research, a non-profit organization in London, England. Throughout its existence, *Mankind Quarterly* has maintained its character as a journal devoted to the interdisciplinary study of man. New developments in the field were reflected in the journal early on. When first sociobiology and then evolutionary psychology developed descriptions of human nature and explanations for cultural universals, the new developments found expression in the journal. The same happened with behavioral genetics, which like sociobiology saw major advances during the last three decades of the 20th century.

Today the editorial board includes scholars from 12 countries who represent a wide variety of disciplines including primatology, physical anthropology, archaeology, cultural anthropology, human genetics, differential psychology, sociology, and history. Despite their varied expertise and views, the editors share a common interest in the evolutionary and historical processes that generate human diversity, and in the universal features of human nature that constrain this diversity. Because history and biological evolution are ongoing processes, this includes an interest in the social, cultural, demographic and biological changes that are taking place in modern societies.

Some of the articles the journal publishes deal with the evolutionary and historic processes that have created the racial, ethnic, linguistic and cultural diversity we see today. Others describe present-day cultural diversity and ongoing trends, especially at the psychological level. During the last years, intelligence and its change over time have been especially active research areas, and a substantial part of this work has been published in *Mankind Quarterly*. The subject is important because education and intelligence are considered the key drivers of cultural change and economic growth, not only in developing countries but also in mature postindustrial societies. Other areas of special interest are the demographic changes that are taking place in countries today and that are shaping the world of the future.

Historically, *Mankind Quarterly* has earned a reputation for publishing articles in controversial areas, including behavioral race differences and the importance of mental ability for individual outcomes and group differences. During the "Bell Curve wars" of the 1990s, it became a target of attack when opponents realized that some of the work cited by Herrnstein and Murray had first been published in *Mankind Quarterly*. However, much of this science has stood the test of time. For example, the importance of genes for individual differences in intelligence is no longer controversial, and genetic effects on individual and group differences that were merely inferred in earlier research are now studied at the molecular level. There is nothing wrong with being at the embattled forefront of new scientific developments.

Most of the research that *Mankind Quarterly* publishes today is "normal" science, but the editors still welcome controversy and new ideas. They see it as part of the journal's mission to provide a forum for the presentation and discussion of theories and empiric research that challenge entrenched beliefs. Of course, the often contradictory views that are represented in *Mankind Quarterly* are those of the individual authors, not those of the journal's publishers or editors.

Notes for Authors

As a peer-reviewed academic journal of anthropology, *Mankind Quarterly* publishes articles on all aspects of the *science of man*, ranging from cultural and physical anthropology, and psychology and behavioral genetics, to demography, mythology and the history of religion.

However, the editors are especially interested in articles relating to cultural and biological evolution, and to the interaction between biology and culture. Such topics include (1) the historical origins of present-day cultural and biological diversity using approaches from history, archaeology, linguistics, mythology, and population genetics; (2) the study of cultural and biological trends in contemporary societies including cross-cultural studies of personality, intelligence and culturally transmitted beliefs and values, as well as the study of demographic trends and trends in physical characters and gene frequencies over time; and (3) the implications of current trends for future human evolution.

Included in these areas of special interest are articles dealing with the evolution of personality, its expression in varying cultural, ecological and economic conditions, and its implications for future cultural and biological evolution. Interdisciplinary approaches that integrate findings from historically separate disciplines and subdisciplines are especially encouraged.

MQ publishes research reports, theoretical articles, data-driven reviews, book reviews, and short communications on new discoveries or critical comments on published papers, either in *MQ* or elsewhere. Because *MQ* is read by individuals from diverse backgrounds, the authors of highly specialized or technical articles are asked to present the background and significance of their work clearly and concisely. Book reviews should also deal with publications which will be of interest to a non-specialist audience.

Theoretical articles may be speculative or controversial or both, but must be based on solid data. Submissions are evaluated only on the basis of scientific soundness, relevance, and interest.

Manuscripts should be composed in **Microsoft Word** and submitted electronically to *submissions@mankindquarterly.org*. The editors may request a printed copy of any article which is accepted for review by the journal's referees. The review process usually takes 1-3 months. Every effort is made to publish articles within six months after final acceptance.

Required typography for manuscripts which have been accepted for publication

Abstracts and Keywords

All manuscripts should include key words, a brief abstract, and an email or postal address for correspondence.

Hyphens

Please do not manually hyphenate words at the ends of lines – not even when preparing tables.

Indents and Tables

Tables must be constructed using the MS Word table-building function.

Kindly do not create columns and indents by pressing the space bar a multiple quantity of times.

Footnotes

Footnotes must be created using the MS Word facility for building footnotes so that the footnote number in the text is digitally linked to the footnote itself.

Footnote numbers must be "superscript", not "raised".

MQ does not use endnotes.

References

References at the end of the article must be alphabetical by author, or, where necessary, name of publication.

References and bibliographic data must be formatted according to the following examples:

Journal articles:

Meisenberg, G. & Lynn, R. (2011). Intelligence: a measure of human capital in nations. *Journal of Social, Political & Economic Studies* 36: 421-454.

Books:

Lynn, R. & Vanhanen, T. (2012). *Intelligence: A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Articles in edited collections:

Rushton, J.-P. (2003). Race differences in *g* and the Jensen effect. In: H. Nyborg (ed): *The Scientific Study of General Intelligence: A Tribute to Arthur R. Jensen*. Amsterdam: Pergamon.

Material published online:

Statistics New Zealand (2013).
http://www.indexmundi.com/new_zealand/demographics_profile.html

INDEX TO VOLUME 57

Book Reviews