genetic distance findings does not permit unequivocal statements about the genetic determination of the present findings. The great poverty of Africa produces conditions that can unquestionably lower IQ such as malnutrition, disease, inferior prenatal care, and inferior perinatal care. The incidence of epilepsy is very high in Africa, and this has been attributed to malnutrition, parasitic and other brain infections, febrile convulsions, and birth injuries (Osantokun et al., 1987).

It could be argued that there are methodological limitations with respect to the mean IQs provided by Lynn and Vanhanen (2002). One possible limitation is that for 74 of the countries used in the present study the IQs were estimated on the basis of IQs in neighboring countries. The present authors maintain that the assumption of neighboring countries tending to have similar IQs is not entirely arbitrary and unjustifiable. We use the analogy of height because height and IQ have similar determinations such as genetics, nutrition, and health care. It would seem more reasonable to predict height of Norwegians from the height of Swedes than from the height of Italians. Furthermore, the fact that the correlations for the calculated IQ countries and estimated IQ countries are similar lends credulence to the legitimacy of the Lynn and Vanhanen procedure for estimating mean IQs. Another possible limitation is that the Lynn and Vanhanen means are based on different tests administered in different eras and in countries that differ in average educational attainment. However, Lynn and Vanhanen made adjustments for the "Flynn effect," an increase in intelligence test performance in recent decades (Flynn, 1987). And, for 50 of the 55 countries in which IQ was calculated and used in the present study, the Raven Progressive Matrices and/or the Cattell Culture Fair Test were used. Although both of these instruments appear to be void of educational and specific culture content, it cannot be assumed that they are equally effective in measuring intelligence around the world. The fact that skin color is not uniform within countries (as displayed in Biasutti, 1967), could also be seen as a methodological limitation. However, the very high inter-rater reliability, combined with the high correlations with skin color, indicate the effects of this limitation are rather small. Furthermore, the positive correlation between skin color and temperature provide evidence for the validity of the skin color map employed. The fact that the *N* for some of the mean IQ's of the countries provided by Lynn and Vanhanen are below optimal constitutes an additional limitation. Measurement instrument limitations ordinarily attenuate rather than inflate correlations. Our correlations are high.

Acknowledgments

We thank Joan Ensher for library research, anthropologist, Roger LaJeunesse for helpful suggestions.

References

Beals, K. L., Smith, C. L., & Dodd, S. M. (1984). Brain size, brain morphology, climate, and time machines. *Current Anthropology*, *25*, 301–328.

Biasutti, R. (1967). *Le Razze e i Popoli Della Terra*. Torino, Italy: Unione pipografiza-Editrice Torinese.

Cavalli-Sforza, L. L., Menzoni, P., & Piazza, A. (1994). *The history and geography of human genes*. Princeton: Princeton University Press.

Diamond, J. (1997). Guns, germs and steel. New York: W.W. Norton & Company.

Flynn, J. R. (1987). Massive IQ gains in 14 nations. *Psychological Bulletin*, *101*, 1171–1191.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Lynn, R. (1991). Race differences in intelligence: A global perspective. *Mankind Quarterly*, *31*, 255–296.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport: Praeger Publishers.

Osantokun, B. O., Adeuja, A. O., Nottidge, V. A., Bademosi, O., Olumide, A., Inge, O., et al. (1987). Prevalence of the epilopsies in Nigerian Africans: A community-based study. *Epilepsia*, *28*, 272–279.

Pearce, E. A., & Smith, C. G. (1998). *Fodor's world weather guide*. New York: Random House.

Rushton, J. P. (1995). *Race, evolution and behavior: A life history perspective*. New Brunswick, NJ: Transaction.

Rushton, J. P. (1997). Race, intelligence, and the brain: The errors and omissions of the revised edition of S.J. Gould's the mismeasure of man. *Personality and Individual Differences*, *23*, 169–180.

Rushton, J. P. (2000). *Race, evolution, and behavior. A life history perspective* (3rd edition). Port Huron: Charles Darwin Research Institute.

Rushton, J. P., & Rushton, W. W. (2003). Brain size, IQ, and racial group differences: Evidence from musculoskeletal traits. *Intelligence*, *31*, 139–155.

Discussions

Comments on correlations of IQ with skin color and geographic–demographic variables

Abstract

A large number of national and geographic population samples were used to test the hypothesis that the variation in mean values of skin color in the diverse populations are consistently correlated with the mean measured or estimated IQs of the various groups, as are some other physical variables,

known as an ecological correlation. Straightforward statistical analyses clearly bear out the hypothesis, showing a significant positive ecological correlation between lightness of mean skin color and mean IQ across different populations. The main limitation of such a study design is that correlations obtained from this type of analysis are completely non-informative regarding any causal or functional connection between individual differences in skin pigmentation and and individual differences in IQ, nor are they informative regarding the causal basis of the correlation, e.g., simple genetic association due to cross-assortative mating for skin color and IQ versus a pleiotropic correlation in which both of the phenotypically distinct but correlated traits are manifested by one and the same gene.

The simple, clear-cut design and statistical analysis in Templer's and Arikawa's (in press) study is entirely conventional, not unlike innumerable other studies published in this and many other refereed psychological and sociological journals. The authors straightforwardly test a hypothesis put forth by other researchers in attempting to explain the observed cognitive racial differences in terms of evolutionary psychology. The main research question is whether a large sample of estimated mean IQs in a great many countries around the world are correlated with the mean temperatures in those regions. And because differences in the darkness of skin color also vary with the mean temperature different climatic regions, the next question is whether skin color, too, is related to IQ. The data are presented, the zero-order correlations calculated, and the hypothesized IQ × skin color correlation is statistically borne out.

The collection of mean IQs and estimated IQs in the various countries (whether *N*-weighted or not) almost certainly fall short of the degree of reliability and validity attainable with psychometric tests administered under laboratory conditions to carefully selected representative national population samples tested under virtual laboratory conditions. Nevertheless the study's rough and ready pick-up mental test data, probably because it was aggregated over some number of varied samples in each country, evidently has sufficient overall precision to show significant and meaningful *ecological correlations* with the other variables in the study. (So-called *ecological correlations* are based on the means of aggregated data, such as the data used in the present study. They are widely used in epidemiology, a branch of medical research, and they also have valuable applications in the behavioral sciences Lubinski & Humphreys, 1996). Ecological correlations are not necessarily less valid than correlations based on individuals' test scores as the unit of measurement, unless it can be successfully argued that, in a given set of data, the nominal ecological variable (e.g., various nations' mean IQs) represents an essentially different latent trait from that represented by test scores based on individuals. It is actually around this important consideration that the crux of the causal argument involving group differences exists between strict environmentalists and most hereditarians, who claim that some proportion of the total between-groups variance is genetic (e.g., Rushton & Jensen, 2005). The aggregation of individual test scores to form group means serves to "average-out" the unique and irrelevant sources of variance among individuals, thereby abstracting and highlighting what is presumably one and the same common factor reflected by the test scores of individuals. This culling out the irrelevant variance unrelated to the latent trait of interest is typically a basic, though not explicit, assumption of most uses of ecological correlations. The use of ecological correlations often calls for more detailed justification than is typically provided in the study of national or other group differences.

My preferred methodology for researching the meaning of correlations between phenotypically distinct variables, such as IQ and skin color, for example, begins with data that is as close as possible to the basic units of its phenotypic measurement in individuals. But to insist on that condition would be to insist that Templer and Arikawa should carry out an altogether different study from the one they have presented here, which is entirely above board in the details of its method, so typical readers of *Intelligence* can make their own subjective evaluations of the study's scientific importance in light of the data's limitations as explicitly announced by the authors. Their analysis yields an interesting set of findings, in the form of several ecological correlations. These could suggest other hypotheses concerning the actual cause of the correlation between IQ and skin color. The cause of the obvious correlation between skin pigmentation and climate (or temperature) is now well understood and is scarcely controversial. It is the relationship between skin color and IQ within as well as between the major racial groups that is of most interest.

Some years ago I reviewed the then total empirical literature on the IQ × skin color correlation (Jensen, 1973, pp. 222–224). The evidence from eighteen published studies indeed leaves very little doubt of a significant correlation *within* the racially hybrid population of African-Americans, in which today about 25% of its genes derive from Caucasian-European ancestors. The overall average of the reported IQ × -

skin color correlations is about +0.20, ranging from +0.12 to +0.30. (The correlations are positive because lighter degrees of pigmentation were assigned higher ratings). These correlations most probably reflect primarily the considerable genetic heterogeneity of a relatively recent hybrid population that has not yet interbred for enough generations to allow all genetic linkages between the genes affecting skin color and the genes affecting cognitive abilities to break up sufficiently to attain genetic equilibrium (i.e., when the Hardy-Weinberg law applies to all of the gene loci involved in the correlated traits). Templer and Arikawa emphasize that they regard skin color only as a climatic variable, a multigenerational reflection of climatic history. And this may well be theoretically adequate for their present purpose. But we should not let it mislead us to dismiss completely other possible, and presently causal, connections between skin color and IQ—an idea the authors, perhaps too cautiously, called "absurd." This stance overlooks the probability of the genetic phenomenon of *pleiotropy* acting as at least a partial cause of the IQ × skin color correlation in present day populations. (Pleiotropy is the condition of a single gene having two or more phenotypically quite different effects. For example a single gene could affect both IQ and skin color.) The main theory invoked to explain the IQ × skin color correlation is that the many millennia of successive migrations out of Africa into Europe and Asia, with their colder and often more severe climates, selected simultaneously both for skin color and for certain cognitive abilities, especially *g*. While both factors were also controlled by separate genes acting independently and would gradually approach Hardy-Weinberg equilibrium, it is also highly likely that in the very long course of this massive selection process even a rare genetic mutation resulting in *pleiotropic* genes that affect *both* pigmentation and *g* would have occurred and then multiplied in frequency because of their joint selective advantage in different climates.

The pleiotropy hypothesis makes sense in terms of evolutionary genetics. But can we empirically reject this pleiotropy hypothesis? After all, the possibility of outright empirical rejection of a hypothesis is the Popperian criterion of scientific argumentation. I do think it is possible to meet this criterion. I propose that it can be done by determining whether the IQ × skin color correlation is what I have elsewhere termed an *intrinsic* correlation, as contrasted with an *extrinsic* correlation (Jensen, 1980; Jensen & Sinha, 1993). The presence of an extrinsic correlation in the absence of an intrinsic correlation rules out pleiotropy. The methodology of making this distinction has been applied to the correlation of IQ with physical stature (extrinsic), of IQ with head size (intrinsic), and of IQ with myopia (intrinsic) (Jensen, 1980; Jensen & Johnson, 1994; Cohn, Cohn, & Jensen, 1988). An *extrinsic* correlation between variables X and Y is one in which the absolute value of $r_{XY}>0$ in the population and is not >0 within families (i.e., within full sibships). An intrinsic correlation is one for which the absolute value of $r_{XY}>0$. both in the *population* and *within* families. Although every pair of full siblings (including DZ twins) has exactly the same unique ancestral genealogy, the members of each pair differ in the particular selection of the parental genes they inherit at conception. An individual who inherits a pleiotropic gene manifests both of its phenotypic effects, such as lighter pigmentation and higher IQ, as would be hypothesized in the case of these two variables. All of the IQ × skin color correlations reported in the literature are entirely population correlations, hence they are not informative regarding pleiotropy. But with a reasonably large sample of full sibling pairs it would be possible to rule out pleiotropy. It would be ruled out if no statistically significant within-families correlation were found between siblings' IQs and the siblings' values on a linear index of skin pigmentation as objectively measured by one of the standard procedures used in physical anthropology. Pleiotropy implies that within each sibling pair the individual having the higher IQ would also more frequently have the lighter skin color. If this is not found to be the case, the pleiotropic hypothesis would have to be rejected. But if, on the other hand, the IQ × skin color correlation turns out to be pleiotropic, and if this result can be adequately replicated, it would constitute a key item of evidence for the co-evolution of IQ (or more specifically *g*) and skin color. Unless geneticists can find sufficient fault with this line of reasoning as to render the proposed study scientifically worthless or technically unfeasible, I would hope that such a study will soon be forthcoming.

References

Cohn, S. J., Cohn, C. M. G., & Jensen, A. R. (1988). Myopia and intelligence: A pleiotropic relationship? *Human Genetics*, *80*, 53–58.

Jensen, A. R. (1973). *Educability and group differences*. London: Methuen (New York: Harper & Row).

Jensen, A. R. (1980). Uses of sibling data in educational and psychological research. *American Educational Research Journal*, *17*, 153–170.

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

Jensen, A. R., & Sinha, S. N. (1993). Physical correlates of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligences* (pp. 139–242). Norwood, NJ: Ablex.

Lubinski, D., & Humphreys, L. G. (1996). Seeing the forest from the trees: When predicting the behavior or status of groups, correlate means. *Psychology, Public Policy, and Law*, *2*, 363–376.
Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *1*, 235–294.
Templer, D. I., & Arikawa, H. (in press). Temperature, skin color, per capita income, and IQ: An international I perspective. *Intelligence*.

Arthur R. Jensen
University of California, Berkeley, United States
E-mail address: nesnejanda@aol.com.

doi:10.1016/j.intell.2005.04.003

Sorry, wrong numbers: An analysis of a study of a correlation between skin color and IQ

Abstract

We argue that the report by Templer and Arikawa contains misleading conclusions and is based upon faulty collection and analysis of data. The report fails to hold up for quality of data, statistical analysis, and the logic of science.

Templer and Arikawa (this issue) report a −0.9 correlation between mean national intelligence test scores and an index of skin color running from 1 (white) to 8 (extremely dark). They interpret this correlation as consistent with arguments by Lynn (1991), Rushton (1995), and Jensen (1998) that both skin color and intelligence are largely biologically determined variables, and that the two covary. Templer and Arikawa further stress an argument, originally due to Rushton, that the greater challenge of living in higher latitudes, plus the paucity of sunlight, favors the evolution of both superior mental capacity and lighter skin color. (One might equally argue that the greater challenge of living in equatorial regions, such as fending off myriad parasitic diseases, should render equatorial people more intelligent.) The authors do admit that there are non-biological explanations, such as Diamond's (1997) argument that the Earth's geography is such that the flow of ideas and technology (and hence prosperity) is easier along the Eurasian East–West axis than the North–South axis, and that societies where there is a constant interchange of ideas are likely to produce more analytic, enquiring minds.

Here we will maintain that the Templer and Arikawa data collection and analyses are seriously flawed. Even if their methods were technically adequate and if the claimed correlations existed, the correlations would be uninterpretable and hence of no scientific value.

Templer and Arikawa's variables are a national intelligence estimate, taken from Lynn and Vanhanen's (2002) analysis of the relation between estimated mean national IQ and gross domestic product per capita (GDPC), an estimate of "preponderant" skin color for the country in question, and mean winter and summer temperatures in the national capital. The data are provided in Templer and Arikawa's Table 1, which also provides the 2001 population of each country. The IQ and GDPC data were taken from Lynn and Vanhanen (2002).

We shall deal with three variables—IQ scores, the skin color index, and population. We shall also distinguish between scores on a putative test of intelligence (throughout, IQ) and intelligence as a concept, viewed here as individual differences in mental competence that can influence a person's success in life.

1. Objections based on the quality of the data

The population data are, of course, no more suspect than any routinely collected census data. The IQ and skin color indices are more suspect.

As Templer and Arikawa accepted the Lynn and Vanhanen data at face value, any weakness in the Lynn and Vanhanen data is inherited by the Templer and Arikawa study. The Lynn and Vanhanen data set is far from ideal.

Lynn and Vanhanen's IQ data were based on reports from a variety of studies in 81 countries. Virtually none of the original studies claimed to be based on national samples. For example, several of their data sets were what the original authors described as standardization samples for the Raven Matrices tests. These standardization samples were by no means population samples. In some cases they consisted of school children in a single town or city. In general, when Lynn and Vanhanen had two or more samples within the same country, they averaged them without weighting for sample size. Even if we accept the validity of the various IQ tests across cultures, a point to which we will return, estimates obtained in this way strike us as being, on statistical grounds alone, inadequate estimates of national IQ.

In order to develop a larger sample of countries, Lynn and Vanhanen estimated IQ scores for a further