



August 9th 2006

Dean Svend Hylleberg,
Aarhus University,
Denmark
Dekan.sam@au.

Dear Dean Hylleberg,

Re: Helmuth Nyborg's Research on Average Sex Differences in IQ

I am writing in support of Professor Helmuth Nyborg, one of the foremost psychologists in Denmark, who has been relieved of his duties following an Evaluation Committee's decision that he failed to show "due diligence" (i.e., was negligent) because he published the results of scientific research which demonstrated that adult men *average* 8.5 IQ points higher than do adult women (based on his calculation of a point-biserial effect size of 0.28). I have read the Reports the University of Aarhus placed on its website and have regretfully come to the inescapable conclusion that Prof. Nyborg has been the victim of political persecution. I am writing this detailed statement in the hope that it will be useful in any future proceedings regarding this case. In doing so, I will provide an assessment of the Committee's decision and place the sex-IQ controversy in historical context.

My Qualifications. I am a professor of psychology at the University of Western Ontario in Canada, a Fellow of the John Simon Guggenheim Memorial Foundation, and a Fellow of several professional societies. I hold two doctorates and have published six books and 250 articles. I conduct research in the same area as Prof. Nyborg, viz., individual and group differences in cognitive ability and in personality and social behavior. Recently, I (Jackson & Rushton, in press) have corroborated Nyborg's finding of a higher mean IQ in adult men (in *Intelligence*, see page proofs at http://psychology.uwo.ca/faculty/rushton_pubs.htm). We analyzed 145 items from the Scholastic Assessment Test (SAT) in 100,000 17- to 18-year-olds and found a male advantage of 3.63 IQ points (based on a point-biserial effect size of 0.12). While lower than Nyborg's estimate, it confirms his general finding. Importantly, our study demonstrated an average sex difference at every level of the SAT, in every level of family

income, for every level of fathers' and of mothers' education, and for each and every one of seven ethnic groups.

Prof. Nyborg's position (and my own) is a minority one in psychology at present, and highly unpopular in many circles, so it is not surprising that some believe that Prof. Nyborg must be a maverick and hope that his data can be made to "disappear." In this context it is vital that justice be "seen to be done," as well as actually done. Since the very act of setting up an outside Committee has so much potential to damage a professor's reputation—as well as that of the university that carries one out—the procedure should only be undertaken when there is a strong and compelling reason to do so. To be seen as fair and legitimate, such a Committee must specify the justification for why an extraordinary procedure is occurring. No such "bill of particulars" has been provided in Prof. Nyborg's case. All that is stated is that the Dean Tom Latrup-Pedersen of the Faculty of Social Sciences decided to appoint a 3-person evaluation committee based on a recommendation by the Psychology Department Head, Jens Mammen, because Regulation 17, subsection 4, of the Statutes Governing Universities, requires the Head to "ensure quality and consistency in the research and teaching carried out at the department," and the "scientific ethical guidelines of the university." This is a very long way from being sufficient justification for such an extraordinary undertaking. Did the Head suspect "fraud" or some other egregious ethical violation? Is there any evidence at all to support such a charge? In the absence of any bill of particulars, the initiation of proceedings appears without foundation and even pejorative.

I have known Prof. Nyborg since the 1983 inaugural meeting of the International Society for Individual Differences (ISSID) in London. I have acquired the highest regard for the integrity of his scientific work, which I have followed closely at meetings of ISSID, including the memorable one he organized in Aarhus in 1997. I have also followed his presentations at meetings of the Behavior Genetics Association (BGA), and the International Society for Intelligence Research (ISIR), including the one in 2001 when discussion of his findings caused a commotion at his university and resulted in the Committee's investigation.

The Committee allows that they could find Prof. Nyborg guilty at three levels of severity—failing to show "*due diligence*" (carelessness); making *unintended mistakes* (negligence), and committing *fraud* (intended mistakes). In the final analysis, the Committee

found a lack of “*due diligence*” in several places and one “*mistake*.” They found no evidence of fraud. Prof. Nyborg may well want to plead “Not Guilty” to even these misdemeanors. Regardless, it is impossible for an outside observer to see how these judgments against Prof. Nyborg could possibly result in such a severe sentence as a suspension of duties and a loss of graduate students. It simply makes no sense except in terms of an organized campaign of political persecution and the “making of an example.”

In order to claim carelessness, the Committee re-analyzed Prof. Nyborg’s data sets in all sorts of different ways. He was forced to supply them with *two filing cabinets containing protocols*, several diskettes of data, and submit to interviews. The Committee’s report is full of summary tables and appendices, and because Prof. Nyborg’s study is a mixed cross-sectional, longitudinal design of 325 8- to 14-year-old boys and girls and 62 adults, with a large drop-out rate and much missing data, and with different tests given to different samples at different times, it is easy to leave an impression of having found “discrepancies” across his several reports in publications and paper presentations. I have looked closely at all the evidence the Committee provided. I found no evidence to sustain any of the charges.

The most egregious errors cited by the Committee include the alleged problem of Prof. Nyborg’s referring to an N of 52 adults in his 2001 ISIR presentation and 2003 book chapter but to an N of 62 in his 2005 journal article. Nyborg apparently explained that in the interim he had entered 10 more cases into the computer, and later that it was simply a typo and should always have been $N = 62$. The Committee re-analyzed all the data and said they were still puzzled because on 15 of the 20 variables they got the same mean with $N = 52$ as with $N = 62$. (They did note, however, that the sex difference remained virtually the same in both cases with a point biserial effect size of 0.272 and 0.280, respectively.) Similarly, the Committee pointed to a slippage of N s with the sample of children, because some analyzes reported $N = 325$ and sometimes $N = 119$. Also, the Committee was puzzled as to how, if all the “children” were now 28-years-old or more, the study could be described as “ongoing,” which the Committee claimed to find “surprising” and “unrealistic.” (Is the child data that relevant to the question of the sex difference in adulthood? Further, I have several studies of my own which I at times refer to as “ongoing” in which data collected over twenty years ago still remain unanalyzed—some of it still in raw data sheets lying at the bottom of filing cabinets! Unlike, “longitudinal” versus

“cross-sectional,” “ongoing” is a term of ordinary discourse rather than of specific definition in the behavioral science lexicon.)

Several pages are devoted by the Committee to trying to replicate Nyborg’s 0.280 *point-biserial factor loading*, which is what he translates into the male advantage of 8.5 IQ points. The Committee reported 12 different analyses and got results ranging from 0.140 to 0.361 (mean = 0.27) but they claim they cannot reach his 0.280 figure “exactly.” (However, one of their 12 analyses does in fact yield a loading of 0.277, which rounds to an “exact” 0.280!) Similarly, they tried to replicate Nyborg’s claimed loading of 0.228 from 119 children using 12 different analyses and reported results ranging from 0.185 to 0.270 (mean = 0.243). They also claim to have uncovered an “unjustified *p* value.” They faulted Nyborg for having allowed his various research assistants and helpers to do his analyses for him without him being completely clear (several years later) as to who had done precisely what. (For the record, I recall that one of my esteemed colleagues carried out an analysis on Prof. Nyborg’s data set when Nyborg visited us at the University of Western Ontario in 2000; my colleague was gratified to find results in favor of higher male scores on the *g* factor since he too had found a similar result.)

The Committee also gave their opinion on whether Nyborg’s conclusion about the male IQ advantage being on the *g* factor was supported by the data. They concluded that it was “indeterminate” because the factor analytic procedure Nyborg used (and Jensen before him), which involved inserting a calculated point-biserial correlation from male-female differences back into a correlation matrix, was a “flawed” procedure and that, in any case, Nyborg had made an “error” in one of his assumptions and so calculated a higher estimate for the sex difference than the Committee would have. The Committee’s calculated effect size using Nyborg’s procedure lay between 0.06 and 0.23 (mid-point of range = 0.15), not the 0.28 that Nyborg had reported. The Committee concluded that Prof. Nyborg should have reported the results from a number of alternate procedures and specified the choices among them in order to estimate the robustness of the phenomena. They also suggested that the inclusion of some of the subtests—such as the Rod and Frame test and a Spatial Rotations test, as well as the way they were coded and, in effect, doubly entered—increased the result in favor of males.

I wish to point out that the Committee’s own best estimate of the point-biserial effect size using procedures currently routine in this literature is 0.15 and this translates into a male

advantage of 4.55 IQ points. In other words, the Committee formed by the University of Aarhus has *confirmed* Professor Nyborg's conclusion that men have a higher *mean* IQ than women. The magnitude of the effect was reduced, but not removed. It would have been more appropriate and stimulating for the Committee to publish an article based on all their re-analyses in the same journal as Nyborg (*PAID*, in 2005) to demonstrate the reduced effect size, as well as discuss general methodological points that may sometimes arise.

The Committee complains that in his 2005 publication Nyborg did not elaborate sufficiently on the various choice points he made in carrying out his factor analysis. However, few journals (or readers) are that interested in such a tedious level of detail. There can always be just one more analysis to be tried, but space in academic journals is typically very limited. Most readers want only enough detail about well known procedures to be able to repeat it themselves. The very top journals in the world such as *Science* and *Nature* have extremely limited space and routinely make the reasonable assumption that researchers are competent to make their own decisions about experimental design and tell their own stories.

The Committee even faults Nyborg for speaking to the media (in 2001) and describing his results in a book chapter (in 2003) before publishing them in full (in 2005). Perhaps this rule can indeed be located in a set of guidelines somewhere in Denmark, but it flies in the face of currently accepted international practice. Surely administrators at the University of Aarhus know that many scientific conferences (including Britain and the European Union) make media rooms available in an attempt to get journalists to take an interest in their presentations. Very few conference papers are given *after* formal publication of the results; in fact, it is often a condition of submission to a conference that the work has *not* already been published.

It is hard for me to reject the hypothesis that the Committee strained to convict Nyborg of *something*, perhaps *anything*, to justify the administrative decision that had set up the Committee initially. It would have been difficult to exonerate him completely without indicting the Head of the Department and the Dean of the Faculty for incompetence and "moral panic." If cooler heads had prevailed and the Committee set up with guidelines specifying a need to "clear the air" and "exonerate the researcher," the impasse would have been avoided. If the Head of the Department had double checked Nyborg's data himself, or had someone go over it informally, he would surely have seen it was sufficiently internally coherent and anchored in the world

literature. It is highly unlikely that *all* editors and referees on so many results from so many laboratories could be so wrong or lax in performing their duties?

Regardless, the administration sought to over-rule the normal process of peer review with no justification provided. Controversial topics such as group differences in IQ are almost guaranteed to get a higher level of scrutiny than usual. Of course, mistakes happen but, given open and free enquiry, science is necessarily self-correcting. Replications and failures to replicate are published and then meta-analytic reviews of effect sizes are undertaken to reveal the “true” state of affairs. The administration at the University of Aarhus appear to have violated these basic rules of science by sending out a signal that “politically incorrect” results will have to bear a very heavy burden of extra proof.

It might be useful for me to review some of the background to the sex-IQ controversy so you can see where Prof. Nyborg’s estimate of an 8.5 IQ point difference falls in the greater scheme of things. For nearly 100 years, ever since IQ tests were invented, there has been a consensus among psychologists that men and women average the same. For decades, however, psychologists have also accepted that men and women differ in their test “profiles,” with males averaging higher on tests of spatial ability and females higher on tests of verbal ability. These differences were assumed to average out. This was the position that Prof. Nyborg held during the 1970s and 1980s (as I did as well).

Two recent sets of observations raised anew the question of sex differences in general intelligence in normal populations. The first was that the general factor of mental ability—*g*—was found to permeate all tests to a greater or lesser extent, as initially proposed in 1904 by the British psychologist Charles E. Spearman. More than any other factor, the magnitude of the test’s *g* loading best determined a test’s power to predict academic achievement and job performance. Thus, a “spatial” test may be relatively high on *g* (mental rotation) or low (perceptual speed), a “verbal” test may be relatively high (reasoning) or low (fluency), as may a “memory” test be high (repeating a series in reverse order) or low (repeating a series in presented order). The question of sex differences in general intelligence thus became formulated more precisely as: “Are there sex differences on the *g* factor?”

The second set of observations concerned the sex difference found in brain size and the relation between brain size and cognitive ability. In 1992, C. Davison Ankney re-analysed 1,000

brain weights at autopsy and discovered that males averaged a larger brain size than females even after adjusting for body size (140-grams before adjustments; 100-grams after adjustments). Ankney's findings were immediately corroborated by Rushton (1992) using cranial capacity data calculated from head size measurements gathered from a sample of 6,000 U.S. military personnel with individual adjustments made for body size. At the time, Ankney's and Rushton's findings were considered "revolutionary" (from an Editorial published in *Nature*) because prior studies of sex differences in brain size argued that they "disappeared" when adjusted for body size (by using an inappropriate adjustment based on brain- to body-size ratios, it turns out). Subsequently, in Denmark, Pakkenberg and Gundersen (1997) documented that men have 15% more neurons than women (22.8 versus 19.3 billion), and over two-dozen MRI studies have confirmed a brain-size/IQ correlation of about 0.40.

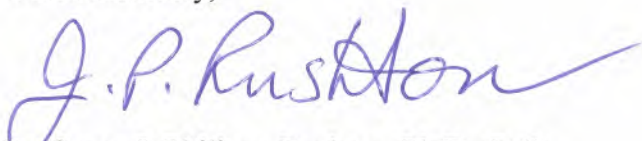
British psychologist Richard Lynn, of the University of Ulster, dubbed the findings on sex differences in brain size "the Ankney-Rushton anomaly." He argued (1994, 1999) that if brain size is linked to IQ, and males average larger brains than females, then men should have higher average intelligence than women. In a series of meta-analytic reviews and new empirical studies, Lynn, along with colleague Paul Irwing at the University of Manchester (e.g., 2004, 2005), showed that on a number of intelligence tests, and in various countries, adult men consistently average 4 to 5 IQ points higher than adult women. (Irwing & Lynn's most recent paper appeared in *Nature* on July 6, 2006.) Lynn also suggested that because girls mature faster than boys, the sex difference is masked during the school years but can be found after maturation, which explained why the sex difference was missed for 100 years. Almost all the data showing no sex differences were gathered on school children.

Not every study and every analysis found the sex difference in general ability. Prof. Roberto Colom, for example, published a negative instance using the standardization of the WAIS-III in Spain. Lynn and Irwing's (2004, 2005) detailed meta-analyses reveal a wide range of effect sizes favoring males, typically from 2 to 8 IQ points. During the 1990s, Prof. Nyborg became involved in the debate. He began to present new analyses of his own and, as we have seen, found an 8.5 IQ point difference. Although this is one of the larger estimates, it is within the normal range. In one paper he also suggested that Prof. Colom had made an error of interpretation in his data and that even the WAIS-III in Spain showed the male advantage.

Although it now appears that Prof. Nyborg's claim of a sex difference in general IQ is correct, this is, of course, not the main point. It should not matter if he proves dead wrong. If academic freedom is to mean anything it must mean that he is free to express thoughts with which others disagree vehemently. Freedom must also mean that he may do so without fear of reprisal in any form. It demands protection from any form of harassment, intimidation, or ostracism by university authorities.

I shall circulate this letter to academic colleagues, including those in academic freedom societies who take an interest in these matters. I look forward to a statement from the University of Aarhus rededicating itself to the principles of academic freedom and free expression and a re-assessment of a personnel decision that results in a greater sense of justice and fair play. This would undoubtedly redound to the benefit of the University of Aarhus.

Yours sincerely,



Professor J. Philippe Rushton, Ph.D., D.Sc.
Department of Psychology,
University of Western Ontario,
London, Ontario, N6A 5C2, Canada
http://psychology.uwo.ca/faculty/rushton_pubs.htm
Tel: 519-661-3685

cc

Rector Lauritz B. Holm-Nielsen, rector@au.dk

Dr. Helmuth Nyborg, helmuthnyborg@msn.com

Society for Academic Freedom and Scholarship (Canada) <http://www.safs.ca/about.html>

National Association of Scholars (USA) <http://www.nas.org/>

Others