45
(se 103)

Er elektronisk
på sagen.
Journalkontoret

Rektor Lauritz B. Holm-Nielsen
Århus Universitet
Nordre Ringgade 1
8000 Århus C

| | |
|---|---|
| AARHUS UNIVERSITET - JOURNALEN | |
| JOUR NR | AU-2005-602-010 |
| MODT. | 30 JAN. 2007 |
| AKT NR. | 57 |
| ANSV. AFD. | LED SAGSB. TBJ |


Forsknings- og
Innovationsstyrelsen
Ministeriet for Videnskab
Teknologi og Udvikling

Danish Agency for Science
Technology and Innovation

Ministry of Science
Technology and Innovation


**Udvalgene vedrørende**
**Videnskabelig Uredelighed**

**Anmodning om eventuelle bemærkninger, Deres j. nr. AU-2005-602-010**

29. januar 2007

Helmuth Nyborg har ved e-mail af 17. januar 2007 afgivet bemærkninger i sagen.

Deres eventuelle bemærkninger i sagen bedes være sekretariatet i hænde **senest den 28. februar 2007.**

UVVU forventer derefter at have tilstrækkelig information til at kunne vurdere sagen.

En kopi af e-mail af 17. januar 2007 vedlægges.

Med venlig hilsen

Annette D. N. Rasmussen
Annette D. N. Rasmussen
Fuldmægtig


**Forsknings- og**
**Innovationsstyrelsen**
Bredgade 40
1260 København K

Telefon 3544 6200
Telefax 3544 6201
E-post fist@fist.dk
Netsted www.fist.dk
Cvr-nr. 1991 8440


**Sagsnr. 612-06-0008**


Ref. Annette Dalsgaard
Nyrup Rasmussen
Telefon 3544 6279
E-post anr@fist.dk


| | |
|---|---|
| AARHUS UNIVERSITET - JOURNALEN | |
| KOPIFORDELING: | LHN, SM, SL |
| UDFØRT 31/1 -07 | AF: BL |

# Annette D. N. Rasmussen

**Fra:** Helmuth Nyborg [helmuthnyborg@msn.com]
**Sendt:** 17. januar 2007 21:54
**Til:** Annette D. N. Rasmussen
**Emne:** Bemærkninger

Noter til UVVU_2.doc    Münchhausen Art_ English.mswor...    Münchhausen Art_ Technical App...

Sekretariatet for
Udvalgene vedrørende
Videnskabelig Uredelighed.

Sag nr. 612-06-0008.
Med henvisning til Deres brev af 06. december 2006 med anmodning om fremsendelse af eventuelle bemærkninger, skal jeg hermed vedhængt fremsende mine bemærkninger.

Med venlig hilsen
Helmuth Nyborg

**Introduction**

Several hundred documents have already been exchanged in this 5-year long so-called Nyborg-Case. Therefore, in order to reduce repetitiveness, I will in this answer refer primarily to my two previously submitted sets of comments to the "Committee for Proper Research" and to comments to Rector. There are, however, a number of general and problematic aspects of the case that I would like to emphasize here, even at the risk of being somewhat repetitive.

**Fraud or framed?**

The most general problem in the case is, as I see it, that it gives the strong impression of an unprecedented personalised year-long highly systematic attempt to pull together a strong case for improper and irresponsible research (framing?), not based on fraud but rather thriving on a couple of minor inconsequential calculation and typographic errors in a huge research project, on a genuine scientific disagreement about an internationally acknowledged and widely used method, on the bare suspicion of a sample bias, and on an entirely misleading accusation for sloppy project management.

Obviously, I stand not alone with this impression. A large number of international specialists in the area - who have had access both to the data and to the final 2005 report, got out of their way and wrote letters to the Rector of the University of Aarhus that to them the case looked like framing and not at all fraud. The three Editors-in-Chef for the respected scientific journal where the report was published, and also the President for "International Society for Intelligence Research, The President for "The International Society for the Study of Individual Differences", and the editor of "Intelligence", wrote letters to Rector to regret that their scientific standards were being questioned by the university. Finally, eight out of ten professor colleagues also wrote to explain to Rector that the distance between premises and conclusion was too large.

**The framing**

First, there is, I believe, no precedence in the history of (Danish) Universities that a director for a scientific institution decides to personally invest five years in a systematic scrutiny of a narrow corner of a single scientist's voluminous research and that without first having consulted with colleagues with special knowledge in that particular area. Very late in the process an "expert" (see later) committee was set up, but only after the scientist had asked for this himself.

It is difficult to understand why the director did not already from the beginning refrain from personally pursuing the case, as the director and the scientist had previously competed for a scientific position (Docentur) that ended in a complaint. The director got the position, but the scientist filed an objection about what he saw as a breach of fairness. The problem was that the director had not then, nor has he today, produced just one peer-reviewed single-authored scientific journal article to be cited in one of the international research registers, such as "PsychoInfo", whereas the scientist already at the time had more than 40 such international articles in respectable journals behind him. The complaint was filed first to the "Fagråd", then to the Dean, then to the Rector, and finally to the Ministry of Education. It is, in my opinion, incorrect that a director does not declare himself incompetent before he personally undertake a 5 year long and painful investigation of a scientist that previously had questioned the director's scientific credentials, even if being formally in charge of watching over the research as director of the institute.

It is worrisome that the director concentrates the full impact of his critique on a very small part of the scientist's research (a sex difference in intelligence), but even more troublesome that it was generalised to disqualify in toto a lifelong research career spanning 35 years resulting in countless articles and books covering a very large number of different topics.

It is worrisome that the director ignores the honoured freedom to choose a particular method. I believe the correct procedure would have been to criticise the method in a professional journal. In this connection it is interesting to see that the director either ignored the fact that the scientist had previously used the very same method in other publications (e.g. Nyborg & Jensen, 2000; 2001a, b) without being criticised, or rather that he deliberately choose to criticise the method only in connection with a report on a highly controversial sex difference. The director admits that he had found himself and the institute under great public pressure after the public had become aware of such a difference. Is this public pressure the real explanation for the ensuing five-year investigation to find at least something wrong with the sex difference project?

The "expert" committee admittedly disagreed with the scientist about the choice of method, but it is worrying to find that the university used this as a vehicle for disciplinary action. (A revisionen of "stillingsstrukturen" stresses that the single researcher can freely chose method (" … den enkelte forsker har forskningsfrihed inden for sit faglige ansættelsesområde med de forpligtelser, der følger af et ansættelsesforhold. Den enkelte forsker har således frihed til at vælge metode, fremgangsmåde og emne inden for universitetets forskningsstrategiske rammer, som fastlagt i udviklingskontrakten."). What is this freedom worth if the selection of a method – right or wrong - elicits disciplinary punishment?

From a legal point of view it is worrisome that the Nyborg-Case has never been clearly defined and explained to Nyborg over these many years. In fact, the case rather consists of a series of different cases, and as soon as one case was addressed properly, new cases just popped up in a never-ending succession. This raises, in my opinion, serious questions about the legality of the whole process.

For example, the first case (January 2002) arose when the director wrote to the scientist that he personally did not believe that there was such a sex difference in intelligence. Apparently, he did not know of the long research tradition (e.g Lynn, 1999, for review), so he wanted to see documentation.

The next case began when the director (incorrectly) claimed that the scientist refused to document the difference. At this point the director informed the press that he looked with gravity on the case. From there on there was no way back - the director either had to prove to the public and to the university that he actually had a grave case or back off and excuse.

The scientist informed the director that it was the journalist (Mona Samir Sørensen from "Politiken") who made up a 27% sex difference (a figure that neither appeared in the approved drafts nor in the data protocols sent to "Politiken"). This information in turn elicited a director's demand that the scientist had to "prove" personally that it was the journalist and not the scientist who had fabricated the difference. The journalist excused her error in writing. When confronted with this evidence the director neither demanded an excuse from the newspaper, nor informed the public about the proper address of the error, nor did he conclude his case against the scientist.

Instead the director confiscated the scientist's raw data protocols reaching 30 years back in time to see if not something suspicious could be found. It is worrisome that the university never informed the public of the fact that there was nothing wrong with the database.

On the contrary, the director called the scientist to a meeting. Neither the dean nor the director informed the scientist before this meeting about the specific agenda, nor was the scientist given the obligatory advice that he better be accompanied by a counsellor, considering that the director had publicly stated that this was a grave case.

At the meeting, the director (and the Dean) required the scientist to produce - out of memory - an exact research protocol for a huge 30-year project involving several hundred people and more than 1200 variables. This information had no bearing on the 2005 research article on sex differences, but an impression of sloppy research could conveniently be established. The university newer bothered to tell the public that a highly precise research protocol was produced after months of hard work. This protocol indicates the exact number of persons examined (graphically as well as numerically) as well as the sex, age, and dates for the visitations. Instead the image of sloppy project management was maintained, handed over to the press, and was eventually incorporated as just another aspect in the unending Nyborg case for later use in terms of disciplinary action.

A new case was raised about whether there were 52 or 62 persons in various analyses, a typographical error that did not change any conclusion.

Further cases were opened with reference to minor, inconsequential, and non-directional errors of calculation (which could all be explained rationally). Other cases had to do with the choice of method; with reference to other non-consequential typos; and with reference to a suspicion of bias in sampling (see above and below).

Feeling being unfairly treated for years and no way out, the scientist chose to discuss some of the scholarly aspects of the case with a number of highly estimated intelligence researcher joining an international intelligence conference in 2005. This provided another springboard for the director and the dean to raise a further case within the Nyborg case, even if the university urges scientists to participate in public discussions. (According to University Law, § 2, stk. 3, universities urge members of the staff to participate in the public debate (***tilskynde medarbejderne til at deltage i den offentlige debat***. (my emphasis)).

It is therefore ironic to see that the university criticises the scientist for addressing professional colleagues in a discussion of the Nyborg Case, because that could make the university appear in a bad light. The scientist had in no way committed fraud and felt framed, and in such a connection it is entirely irrelevant whether it was the "Committee for Proper Research" or the director that had reprimanded the scientist!

With respect to addressing the press, the truth is that it spares no means to make a sensation out of even a modest sex difference. It was the press that beleaguered the scientist – not the scientist that senselessly exposed himself in order to get at least some attention. The university, of course, knows that this puts an enormous pressure on researchers for them to provide a "good" story, and also knows that the press often twists the information given in sometimes unpredictable directions, as was the case here. However, instead of helping the scientist out of this, the university began to systematically scan the press (and the internet!) in search of possible misunderstandings or sign of disloyalty, and then used whatever it found against the scientist.

It is deeply worrisome to find that the university in fact uses a grossly biased and polemic newspaper chronicle as one of its main pillars in the never-ending intent to discredit the scientist. In that chronicle Pia Ankersen et al. (2003) accuse the scientist of fabricating a sex difference that is not in the data by using a method aimed for this (Hierarchical factor analysis). It is easily demonstrated empirically that this claim is entirely untrustworthy. The university nevertheless choose to incorporate the article as an important act in the case against the scientist.

It is food for thought that neither the director nor the university seem to have ever closely examined the article for trustworthiness, in conspicuous contrast to how they treated everything else pertaining to the scientist. On the contrary, the university pointed to its bounden duty to act with force upon serious collegial critique. When the scientist asked the university to identify the author(s) of this alleged serious collegial critique, it responded by providing two names (Professor Doreen Kimura and Science Magazine senior writer Constance Holden). Those names are entirely irrelevant for the question. Only after the scientist insisted on learning about the correct identity for the collegial critique did the university eventually release three names. Then it became obvious that it is the questionable newspaper chronicle by Ankersen et al. that the university used as one of its main pillars in its scientific case against the scientist.

It is equally worrisome that Pia Ankersen was easily able to convince the University that the scientist had refused to cooperate and to provide documentation about a year after the scientist had, in fact, provided Ankersen with ample documentation. With this documentation in hand, the scientist invited Ankersen no less than four times to comment on the material or to initiate a collegial cooperation, but without ever getting an answer to any of these invitations.

Being entirely unimpressed by this, that university instead focused on the status of some 2001 conference notes - which Ankersen claimed was a secret paper - even years after she was being informed that there was no paper only but conference notes plus tables and figures, to be written up and published later.

It was comforting that the then Rector Chr. Sidenius in a public statement admitted that the scientist must be given sufficient time to write up the final report, but also typical for the case in general that the "Praksisudvalg" nevertheless recommended that the scientist be reprimanded for refusing to deliver the documentation at what they considered the right time. It has not been possible to obtain a reference to the formal rule defining exactly what is this right time for a publication, or whether a scientist is allowed to speak to the public after a conference presentation.

A recent study by Wicherts et al. (2006) is interesting in connection with the patently false claim that the scientist had flatly refused to (ever?) provide data for reanalysis. The study illustrates the common international practice when scientists are asked to provide data for reanalysis. The Wicherts-group wanted to obtain access to data sets for every article published during the last two 2004 issues of four eminent major APA Journals (with rejection rates of over 70%) in order to assess the robustness of the data. The authors had all "...signed the APA Certification of Compliance with APA Ethical Principles, which included the principle on sharing data for reanalysis. The principle is as follows: After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through re-analysis ....".

After considerable effort they succeeded in obtaining actual data sets from only 25.7% of the total number of data sets. This means that the large majority of international researchers are either unable or unwilling to share their data even after having signed an official agreement to do so.

I obviously do not take this to mean that scientists should withhold their data. Rather, the study illustrates that it is internationally quite unusual that scientists actually provide their data even when asked to do so and after having stated in writing that they are willing to do so. Needless to say that none of the authors of the several hundred studies were persecuted for withholding data. Apparently, in the present case the University of Aarhus maintains a much higher standard than do the rest of the scientific world.

Moreover, it is equally obvious that the Nyborg Case also differ in two other vital respects from the Wicherts studies. First, I promised in no unclear words and right from the beginning (in writing, verbally, and in the press) that anybody who cared to ask would be handed over all the relevant material when the report explaining the intricacies of the study was finish. The reason for not immediately giving away the conference tables and figures to everybody was that when unaccompanied by a report explaining the meaning of the number, they had already created havoc and serious misunderstanding when handed over to the press (read: "Politiken"). As mentioned above, Rector Niels Chr. Sidenius actually granted me in a public statement that such a delay was an acceptable solution. Second, despite this the director and the university ("Praksisudvalg") did not hesitate to open another case within the Nyborg case, and instilled that the scientist was given a reprimand. Typically, all this was later incorporated for further disciplinary actions.

With respect to the minor errors found in the Nyborg study, it is enlightening to refer to a personal communication (23. October 2006) in which dr. Jelte Wicherts wrote to me: "You might be interested to know that in about half of the data sets we did receiver, there was something wrong with the data" and furthermore that "... this result also shows that **if you want, you will always be able to find that something is wrong with the data**. Simulation studies have shown that such violations do generally not effect the conclusions (e.g., type 1 error rate) to a strong degree." (my emphasis).

This is a rather precise prescription for how to best frame Nyborg in the Nyborg case or, for that sake, any scientist. There were admittedly errors in my analyses but they were minor and did in no way affect the main conclusion: The expert committee confirmed that it was able to reproduce my results qualitatively. On is reminded of Niels Bohr's definition of an expert as one who has made all the errors.

All of the above suggest that the university really wanted to find something wrong with the observation of a sex difference in intelligence, so that the director could prove that he, in fact, had the serious case he told the press he had. It took the director five years of painstaking search and the turning of every stone, to find only minor errors in a very complicated process involving hundreds of calculations and decisions regarding analytic strategy. The non-directional errors found are understandable consequences of intricate research projects. Errors are, obviously, always to be avoided but, as illustrated above, few international researchers entirely escape this danger. It is this common knowledge that allowed some of the eminent specialists in statistics and research methodology to guarantee (as mentioned in their protest letters to Rector; previously submitted to UVVU) that they will certainly be able to find errors in all the research projects done at the University of Aarhus (or elsewhere), in particular when given enough time (5 years?) for scrutiny and sufficient economic resources. That being the case one scientist, obviously, is no match against the whole brutal machinery of a determined large university.

The inescapable conclusion is that the university unfairly singled out a particular (controversial) projects (out of hundred other projects) of a particular scientist (out of the works of thousands of scientists). After a long and intense search it proved able to find only a couple of non-directional errors, but not the slightest evidence of fraud; it expressed a disagreement about method and aired an unproven suspicion of sample bias (see below). From this it resolved that the scientist should have known about the "many serious deficits" of the study, about secret assumptions, about the use of wrong methods, and that the scientist had demonstrated an obvious lack of statistical insight. The university reasoned that all this led up to a totally worthless and untrustworthy paper on a sex difference in general intelligence, even after it had been published by a respectable peer-reviewed research journal characterising it as eminent, and even after the university had received fierce protests from all three Editors-in-Chief for "Personality and Individual Differences", and from the Editor-in-Chief of the expert journal "Intelligence" by professor Douglas Detterman.

As testified by the many support letters and by the letter to Rector from 8 out of 10 professors in psychology at the Department of Psychology, most commentators regret this and express rather frankly and in no unclear terms the opinion that something else must be at stake here.

**Relevant scientific production**

Rector emphasizes, in his letter dated 3. November 2006, that the members of the expert committee have a ***relevant*** scientific production (my emphasis). I have not previously questioned this, but the remark made me look into the matter as some of the conclusions the committee reached surprised me.

Interestingly enough, the two statisticians have neither personal nor co-authored relevant works on intelligence research, as far as can be judged from their homepages. This may explain why they find it a serious omission that a scientist does not account for the outcome of all possible factor-analytic possibilities in order to control for robustness of result. Apparently they do not know that it long ago was demonstrated empirically (and convincingly), and duly reported in the research literature that various factorial analyses provide qualitatively similar results (for details, see the previously letter of comments to the 10 critical points in the Dean's report, or Jensen, 1998, p. 73). Apparently, the two statisticians do not know that for exactly this reason the editors of the relevant professional journals refuse to use costly space on reporting the obvious. Ironically, the committee's own application of many different factor solutions to my data confirms this point, as previously detailed.

With respect to the only professional psychologist on the expert committee it is obvious that he does not represent mainstream intelligence research, but rather entertains a deviant standpoint in the literature. Herrnstein and Murray (1994, p. 770) thus mention in connection with a basic discussion of bias in the calculation of Spearman's *g*-factor and the investigation of Spearman's hypothesis that "The data confirming Spearman's hypothesis ... provide the most convincing conceptual refutation of this allegation ... borne out by many studies .... Note 12: For a review of the literature throughout the early 1980s, see Jensen 1985 ... for studies since then see Braden 1989; Jensen 1992; 1983. **THE SINGLE CONTRARY STUDY EXTENT IS GUSTAVSSON 1992.**" (my emphasis).

Moreover, the expert committee regrets that it could not reproduce the results exactly. This surprises no statistician with expert knowledge in factorial intelligence research (see for example the protest letter from prof. Feinberg; previously submitted to UVVU). Truth is that the second or third order *g* factor is a latent variable that cannot be measured directly (Jensen, 1998) but has to be estimated by analysis of a correlations matrix, and the number of permutations is very large. Second, the committee demonstrated that the mean of the their many re-analyses amounted to 0.277, a figure that must be said to be satisfyingly close to the 0.272 reported in the final 2005 article, if not exact.

It is rather problematic that Rector on the basis of a Google Scholars search chooses to emphasize the particular relevance of the scientific production of the members of the expert committee, when two out of three members have no documented insight in said research area. However, it explains their appalling lack of insight into traditional intelligence research in a matter of great importance for their critique of a research publication. Add to this that the third member of the committee stands alone in the international research literature with his serious critique of a widely used approach that I also applied, and the terminology "particular relevance" becomes peculiar.

Rector further refers in his letter of 3. November 2006 to the often-mentioned bias problem which is supposed to render my sex research totally worthless and untrustworthy. First, all selection of subjects introduces a bias! Second, the present study originally drew children from primary schools (where bias is considerably less serious for this kind of study than if drafting other subjects (Jensen, 1998)), and the study used a very specific and detailed approach to secure a fairly representative sample - a fact never mentioned by the university in connection with its sample bias critique. Third, under bullit 4 Rector refers to the very considerable drop out problem ("...et meget betragteligt drop out problem ...", but this problem needs more differentiation than just that. Thus, about half of the subjects came from a cross-sectional part of the main study where the participation rate was 100%. The other half came from several longitudinal streaks of the main study. Here the drop out rate differed by group partly as a function of which part they belonged to. For example, the drop out rate was much higher in the control group that in the other two longitudinal groups, as the director repeatedly refused to finance re-examinations in the project. The speculation that perhaps repeated blood sampling and nude photography had contributed to the drop out rate in the longitudinal parts, with an effect on cognitive testing, is definitely countered by the fact that the cross-sectional and the longitudinal children obtained similar cognitive scores (as previously reported.). It is rather depressing to find that the university found no reason to tell the public about this when it was informed about the result of the comparison. Moreover, the university has since 1981 been fully informed about the blood sampling and nude photography, and about the acknowledgement of the relevant ethics committees, the children, the parents, and the schools. In fact, the university has partly financed the collection of such data. It therefore cannot come as a full surprise that these were essential parts of the study. That the director personally also knew about this can be seen from the fact that he provided Rector with a detailed overview of all variables in the main study years ago.

Obviously, whatever can be used against the project will also used against it, even if in an unfair or misleading way. For example, the report from the expert committee states that my research method cannot indicate whether the sex difference is on the *g* factor or in the first order factors. This "bad news" is readily mediated to the press, but neither the committee nor the director nor the university care to inform the public that I also applied a correlated vector analysis indicating that the sex difference is, in fact, a significant *g* effect. They rather concentrate on stressing that another significance test was used incorrectly. The university did not hesitate to expose my appalling lack of statistical insight (at about the level of ninth class) to the public in connection with sex difference research, but never mentioned that since 1971 I have applied – and I might add: with some success - just about all known statistical tools in countless international peer-reviewed publications. It is equally depressing to find that the university's selective lack of ability of provide vital pieces of positive information about the project and about how well it has been managed over thirty years seems carefully designed to furnish the expert committee and the public with an impression of gross neglect.

## Concluding remarks

The list of cases within the Nyborg-Case is much longer. However, the above list suffices to document that there is not a single well-defined Nyborg case. The case rather expands into an endless series of highly systematic yearlong attempts to find at least something seriously wrong with the data, with the method, with the selection, with the interpretation, or with the project in general, or to selectively omit positive elements. It is worth noting that even after this painful process took place, no serious or deliberate faults were found. Instead, the committee was able to qualitatively reproduce the results, and even found that the material was reasonably well defined except for an inconsequential typing error of 52 versus 62 subjects.

This is no less than a remarkable outcome. This is, indeed, a very long way from the accusation for scientific fraud by deliberately or grossly negligent behaviour in terms of falsification, plagiarism, or suppression resulting in improper misleading research. The expert committee was unable to find secret construction of data or substations with fictive data; it could not find secret selective or hidden rejection of own unwanted data; it perhaps suspected unusual or misleading use of statistical methods, but the international elite disagrees with them in this matter; it could not find secret one-sided or twisted interpretation of results or conclusions; and it could not find traces of plagiarism.

What the university found after these five long years is thus a far cry from scientific fraud. This fact was duly noted by many of the internationally most eminent researchers in the field, who had access to the data and to the critique, and therefore sent highly critical letters to the university, as did eight out of 10 professor colleagues sending a collective letter of complaint to the university.

Moreover, many of these international researchers found it deeply problematic that general methodological questions had now become a matter of decision by local committees and bureaucrats at the University of Aarhus. It is indeed a rather dangerous path to take to ask local committees with a narrow commission to post hoc criticise peer-reviewed articles already published in professional journals with the explicit intention to elicit local disciplinary punishment of single scientists. Many of the support letters reflect that most experts in the field see such procedures as more or less hidden attempts by the university to intimidate an individual researcher working in controversial areas, and as a threat to free speech and choice of method in academia. Critique of particular studies or methods belongs properly, in the opinion of most leading scientists, in the professional literature and to knowledgeable research fora.

In short, the Nyborg Case(s) looks, in the opinion of the experts, more like a systematic personalised attempt to use all available means in order to frame for any price a scientist who dared to publish on a controversial topic. It certainly does not look like a case of gross neglect.

Sincerely yours

Helmuth Nyborg, Prof. Dr. phil.

**Previously submitted to UVVU:**

Two sets of comments to the expert committee.

Comments to Rector.

Letters from international editors, presidents, and other experts

**Enclosed:**

Ankersen, P.V., Poulsen, J.J. & Kristoffersen, S.B. (2003). Münchhausen i Århus, Jyllands-Posten, JP-Kronik, torsdag den 9. oktober (English translation).

Ankersen, P.V., Poulsen, J.J. & Kristoffersen, S.B. (2003). Technical appendix to the chronicle in Jyllands-Posten, torsdag den 9. October (English translation)..

## References


Braden, J.P. (1989). Fact or artifact? An empirical test of Spearman's hypothesis. *Intelligence, 13*, 149-155.

Gustavsson, J.-E. (1992). The relevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27*, 239-247.

Herrnstein, R.J. & Murray, C (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press (p. 770).

Jensen, A.R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *The Behavioral and Brain Sciences, 8*, 193-258.

Jensen, A.R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research, 27*, 225-233.

Jensen, A.R. (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence, 17*, 44-77.

Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westwood, CT.: Praeger.

Lynn, R. (1999). Sex difference in intelligence and brain size: A developmental theory. *Intelligence, 27, 1*, 1-12.

Nyborg, H. & Jensen, A. R. (2000a). The Black-White Differences on various psychometric tests: Spearman's hypothesis tested on American armed services veterans. *Personality and Individual Differences, 28*, 593-599.

Nyborg, H. & Jensen, A. R. (2000b). Testosterone Levels as Modifiers of Psychometric g. *Personality and Individual Differences, 28,* 601-607.

Nyborg, H. & Jensen, A. R. (2001). Occupation and Income Related to Psychometri g. *Intelligence, 29,* 45-55.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006, October). The poor availability of psychological research data for reanalysis. *American Psychologist,* 726-728.

Translation from the Danish of the newspaper article:
Münchhausen I Århus, JP-Kronik, torsdag 9.10.2003.
Jyllands-Posten, Århus, Denmark,
Af Pia Vedel Ankersen, Ph.d. stipendiat i statskundskab, Jørgen J. Poulsen, lektor i statskundskab, og Siggi Brandt Kristoffersen, Ph.d. stipendiat i statskundskab, alle Aarhus Universitet.
Münchhausen In Århus
We have not - since Münchhausen escaped the moor by pulling himself up by his own hair - seen a similar trick, writes the chroniclers of the day about professor Helmuth Nyborg's conclusion methods in a research project based on a secret article he will not release.
The article is written by Pia Vedel Ankersen, Jørgen J. Poulseen, and Siggi Brandt Kristoffersen, all from Political Science, at the Faculty of the Social Sciences at the University of Aarhus in Denmark.

After the nuclear war has extinguished almost all life on earth the few surviving amoebas are laying in the dark and thinking about the situation. Then they take a decision: Next time there will be no brain development. Well, perhaps a rather simplified and one-dimensional view on the case.
Of course, there is a large difference between Helmuth Nyborg and the amoebas. But the professor has an equally simplified even if contrary understanding of the consequences of intelligence. He can measure "general ability", the g-factor, and he seems to think that the ability to be a constructive citizen depends on the g-score. Therefore it is perhaps sensible enough to let the pay-cheque for children depend on the g-score of the parents.
Before we look on Nyborg's newest trick, there can be a reason to go back to last year's Nyborg tale about the lower intelligence of the females. We can now say something more about the basis for the debate taking place then. Now we have access to an article where Nyborg refers to his own research results and put them into perspective by comparing them to the results of other researchers.
But we still miss the original research report, where Nyborg presents the documentation for his sensational conclusions.
That article he would not hand over last year and neither will he give it to us this year, even if he refers readers of the new article back to the old secret article if they want to learn about the details of the analysis.
A researcher who presents a new study for his scientific colleagues normally will strive to document how the conclusions follow from the premises.
In that way we get – expressed popularly – a view into the inter-calculations. But that is not what Nyborg will give us. He have, however, given us enough to say for sure that his data cannot carry though the conclusion that women are, on average, less gifted than men.
The entire basis for this conclusion is an analysis of 52 persons. From the information that Nyborg has chosen to hand over to the public we can conclude that the claimed average difference in intelligence between men and women is statistically speaking non-significant. That is to say that the difference that is not particularly large could appear by coincidence. Obviously, Nyborg himself has figured that out, so he does not report that sad fact.
Instead he becomes very creative. He simply uses sex as a co-determining factor in the calculations of his beloved g-factor. In that way he automatically gets a larger, even if artificial connection between sex and intelligence. It is a bit like the old story about the butcher who presses down with his thumb before he decides how much meat is on the weight. After that trick Nyborg feels that he can ignore the real significance test that gave the wrong result.
We have not - since Münchhausen saved himself out of the moor by pulling his own hair - seen a similar trick.
It is instructive to se how Nyborg treats his big hero the American intelligence researcher Arthur Jensen. Jensen finds no sex difference with respect to the g-factor in his study. Nyborg reports this and launch the following critique of Jensen. You can measure the g-factor in many ways and the result you get is entirely dependent on the test that you have included in the measurements. It is known that there are some tests favouring women whereas others favour men. Jensen comes to the wrong conclusion because he has used to many tests that favour women.
This is no less than an amazing admission because in this light the g-factor appears as a purely social construction. You can chop it until it fits the conclusion that is to be gained.
But how does Nyborg then figures out which balance there should be between tests that favour females and tests that favour men? Nyborg here sees life as one big intelligence test.
As there are not as many women as men in high status positions then women are less gifted than men. This must be taken into account when you screw your g-factor together.

Our guess is that Nyborg – despite his disappointing results – has squeezed quite a lot of tests into his analysis that are known to favour males.
We cannot know this for sure because the details are in the secret article. But the conclusion of it all is that because women are underrepresented in high-status jobs we have to construct our intelligence test in such a way that women on average score lower than men. Afterwards we can use the low g-factor to explain the underrepresentation of females. This is a very nice round logical circle which Nyborg has drawn here.
We think that readers at this point will agree that if what we have claimed here is correct then it will be very stupid to base serious politics on the g-factor. But how come that Nyborg has gone so much awry?
Here we have to look a bit closer on the very idea of measuring general intelligence as a one-dimensional phenomenon. We quite agree with Nyborg that it is sensible to see intelligence as the ability to treat complexity.
The problem is, however, that our world offers several different forms of complexity. Nyborg has focussed his conception on mathematical-analytical intelligence. But he also realises that there are other forms of complexity calling for different types of intelligence.
The most obvious defect in Nyborg's g-factor is that it does not take into account personal and social intelligence. Personal intelligence is the ability to be in touch with your feelings and to see oneself from the outside. Social intelligence is the ability to take other people's perspective and put oneself in a reasonable and sensible way in connection to that.
These two types of intelligence, which often are knit together in the concept of emotional intelligence, are of decisive importance for our ability to deal with social complexity.
Emotional intelligence is of great importance for the ability to form constructive and cohesive relationships with other people.
Emotional intelligence is important for the harmonic self-understanding of an individual. If there is one thing we should wish for our children it is precisely that they are becoming socially competent and harmonic people.
It would be good for society and it would be good for them. The fact that the g-factor does not hang together with emotional intelligence is a decisive reason why Nyborg's project is scary. A narrow-minded promotion of Nyborg's g-factor risks to create a society of more autistic individuals.
Nyborg admits that emotional intelligence is important. But he cannot measure it quite well and it cannot at all be measured in a way that makes it hang together with the beloved g-factor. Therefore Nyborg narrows his argument down to the g-factor.
This argumentation is fragile, logically speaking. Suppose that professor Nyborg walks in a dark street with light only from a lamppost some distance away. He walks and plays with his purse and suddenly he looses it. If Nyborg used the same method to find his purse as in his research to study intelligence then he will go forward and look for his purse under the light instead of looking for it in the area where he has lost it.
With such a strategy one gets poor in the long run. We will all be poor if people suddenly began to take Nyborg's ideas seriously.
Gottfried, August Bürger (1980). Friherre von Münchhasens vidunderlige rejser og eventyr til lands og til vands sådan som han selv plejede at fortælle dem ved et godt glas i sine venners kreds. Hernov.
Nyborg, Helmuth (2003). "Sex differences in g" in Helmuth Nyborg (ed.) The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen. Oxford: Elsevier Science Ltd., pp. 187-222.

**Technical appendix to the chronicle in JP, 9. October 2003**

By

*Pia Vedel Ankersen, ph.d.-stipendiate in political science, Jørgen Johannes Poulsen, lektor in political science, and Siggi Brandt Kristoffersen, ph.d.-stipendiat in political science, all at Århus University.*

In the JP chronicle of 9. October 2003 [we] presented two claims of serious faults in professor Helmuth Nyborg's (HNs) investigation of the intelligence of men and women. The purpose of this appendix is to spell out the errors in a more technical way than is possible in a chronicle.

The treatment of the presented results is based exclusively on the article [chapter]: Nyborg, Helmuth (2003). 'Sex differences in *g*.' in Helmuth Nyborg (ed.): *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen.* Oxford: Elsevier Science Ltd., pp. 187- 222. Here HN refers primarily to results presented in a paper from 2001. This means that it is difficult to find out to which extent HN complies to common research criteria, if one reads only this newest publication. Because HN despite repeated requests has refused to make his data or the central paper from 2001 public, more of the claims presented will have to be based on his interpretation of his own results.

Despite this it is quite obvious that there are clear-cut problems with the conclusions he draws. In the following we argue that the much debated sex difference in male favour is in reality associated with such a large degree of uncertainty than HN could equally well have written about the lack of a difference among women and men or even about the superiority of the intelligence of the women. [We argue], moreover, that he – consciously or unconsciously – makes fundamental errors in the analytic approach in a desperate attempt to maintain the advantage of the men.

## The sex difference

In the above mentioned book from 2003 HN gives us for the hard numbers to prove the higher male average score (p. 214). On the famous *g* factor men obtained on average a score of 0,36, whereas women only scored -0,01.[1] Obviously, it is important here to take into account that we here do not

[1] The attentive reader notes that the average g-score for the 52 respondents cannot be zero which would be the normale average value of a factor. This is because the *g*-factor is derived in a larger group of respondents (n=376), who collective

talk about the total Danish population, but only a minor fraction. In the present case actually a very small fraction of 52 individuals aged 16+ years. Like for all other studies with the ambition to speak in general on behalf of the population as such, HN must therefore establish an interval of confidence around the size of the sex difference. Despite the fact that he nowhere presents the result of this, it can be seen that he nevertheless chooses to claim validity of the reported difference. As the following analysis indicated, this is not in conformity with the ordinary rules for statistical inference.

On basis of the information HN gives us, the margin of confidence can be calculated according to the following formula:[2]

$$\text{Sex difference} \pm 1{,}96 * \sqrt{\frac{s^2_{men}}{n_{men}} + \frac{s^2_{women}}{n_{women}}} \quad \longleftrightarrow \quad 0{,}37 \pm 1{,}96 * \sqrt{\frac{1{,}06^2}{31} + \frac{0{,}74^2}{21}} \quad \longleftrightarrow \quad (0{,}37 \pm 0{,}49)\ g$$

The outcome of the calculation is thus that the male advantage in intelligence, expressed by g, lies within the interval of -0,12 to 0,86 *g*. This equals that the average male in an IQ-test scores somewhere between 1,8 point *less* and 12,9 point more than the average woman. It thus cannot be rejected the male in fact may have a higher *g*-score, but the opposite could also be the case. Under normal circumstances, the only statistically tenable would accordingly be to conclude that no reliable sex difference has been found. A direct statistical test for whether we can be sure that the sex difference does not equal zero gives a p-values of 0,139.[3] This would not ordinarily lead to accept of the existence of the sex difference, as it means that 13,9 % of all reported scientific results were directly wrong. With respect to intelligence research, where large disagreement exists about these connection and where the consequence of reporting wrong results can be very comprehensive, it may be reasonable to argue that the common accept of errors at the 5 % level is set too high. In particular when one advocates for such comprehensive societal consequences with background in research results that HN several times has advocated. Therefore, one could argue for an even broader interval then the one calculated about, which would further invalidate HN's result.

---

have an average of 0 *g*. As a sex difference could be found exclusively among the 52 participating adults (at least 16 years old) it is this group only that is used for further analysis.

[2] A confidence interval of 95% is used here, which is standard for the majority of social science research. Helmuth Nyborg does not inform what the sex distribution of the 52 respondents is, just that they are evenly distributed.. As a distribution with 31 men and 21 women is to his advantage in the calculation, we take this to be our point of departure. However, because it is more likely that Nyborg uses an equal sex distribution vi have also performed the analysis based on this assumption. The confidence interval shows in this case that men score from 1,9 points less to 13,0 points more then women in an IQ test (expressed in terms of *g* the interval spreads from -0,13 to 0,87 points).

[3] Assuming an equal sex distribution the p-værdien equals 0,144.

It can accordingly be concluded that ether does Helmuth Nyborg not tell that he with an ordinary level of confidence disconfirm his central hypothesis, or he has deliberately lowered the criterias for a scientific proof to such an extent that he gets the result he wants. Neither of these possibilities speaks in his favour.

**The rescue**

Perhaps in silent acknowledgement of the disappointing results, HN realises that a "simple" test like the one above is not sufficient to reveal a true sex difference. Instead is used an approach which by the power of a particularly exotic Schmid-Leiman transformation and the missing account of the principles behind the analytic method throws the reader into either a higher degree of confusion – and that was perhaps the intentions – or a submissive accept that HN must know what he is talking about. Whereas it from only the description in the 2003-book *is* difficult to find out precisely what goes on in the analysis, it is obvious that the calculation of the *g*-factor suddenly changes. Previously the hierarchical factor analysis was used which quite sensibly weighted the different types of tests – and by that different forms of intelligence – equally in the analysis, but then sex is introduced as an independent variable in the factorial deduction of *g*. In continuation of this HN then thinks that one can decide in the matter of the sex difference by looking at the contribution of sex (*loading*) to the merged factor (p. 209). However, this is like cheating with the weight; if sex contributes as one of the part elements that determines the individual's *g*-score, one cannot afterwards independently test whether sex makes a contribution to the *g*-factor – as one has already decided that it will. One can compare this approach to an intelligence test, where all men are give bonus points for their sex, and then investigate whether the sex of the subjects is important for their score.

After taking this little helping hand to the men into consideration it is sensational that the result only show a difference in their favour that borders up to what normally is seen as statistically tenable ($p \approx 0{,}052$). For this reason HN decides a priori that theoretically it only makes sense if a potential difference in intelligence would favour men, by which the level of significance [usikkerhed] of a possible difference can smartly be halved ($p=0{,}026$). This way of proceeding fits in many ways well to the general picture of his analyses which, as their most important validity criterion, have to support the conclusion which seems written from the outset: That men ought to occupy the most influential positions in society, because they verifiably are the fittest due to their superior intelligence.