Comments to the report by "sagkyndigt udvalg til bedømmelse af Helmuth Nyborgs forskningsprojekt vedrørende kønsforskelle I intelligens"

By Helmuth Nyborg

1.1.1 The nature and quality of the data sets

- The explanation for the discrepancy between N = 52 and N = 62 is that a typing error of 52 made in 2001 was corrected to 62 when the data set later was fully updated. This typing error had no influence on the results. The dropout problem is common to all longitudinal studies. As a rule intense long-lasting studies suffer a dropout rate of 50 percent or more. There is no indication that this had a biasing effect on the cognitive variables. Moreover, as indicated in appendix 1, attrition in the present study does not differ much from the expected attrition in general (also see 1.2.1.4.). In my 2005 publication the reader is, in fact, informed that subjects are drawn from a total study – part of them recruited from cross-sectional and the other part coming from a longitudinal streak of repeatedly tested subjects. This neither raised questions from the reviewers nor from any professional readers, even if they obviously all are well aware of the common drop-out problem in longitudinal analyses. Nobody, but the committee, find my mentioning the longitudinal recruitment of sufficient relevance to raise questions about it in the present context.

- The sample of 325 observations is irrelevant for evaluation of the adult sex difference. It was used at a conference in 2002 to present and discuss various methodological problems and was in passing mentioned elsewhere but not in the final report on the adult sex difference. This means that the critique of mean substitution etc. is not at all relevant for evaluation of the adult sex difference and should not be confused with it, at the committee does.

1.1.2 Documentation and access to reports and data

First, even if it is claimed many times, it is definitely not true that I have ever refused access to the documentation (see 1.2.1.6). Second, I promised in writing and verbally (and this can be documented) that all interested parties would get the full report when it was finalized. Third, on page 10 the committee mentions that Pia Ankersen lodges a complaint that she has not been able to obtain Nyborg 2001 from Nyborg. What she neither tells the committee nor anybody else is that I sent all the necessary information to her, including tables and figures, about a year before she made her complaint. Neither does Ankersen tell the committee nor the press that I sent no less than four emails, repeatedly urging her for a discussion of my data and chapter sent to her earlier on. She never replied to this but nevertheless tells the "Committee for Good Scientific Practice" at Aarhus University and the public press that I flatly refused to cooperate. Fourth, the committee notes that I made the study public in an interview in "Politiken" while denying access to the documentation. This is also not at all correct. "Politiken" got the interview only AFTER the study was presented at the 2001 conference and then "Politiken" got the 2001 tables and figures (also see 1.2.2.6). Fifth, the committee mentions on page 10 that the president of the university took no action. Actually he did. He stated in public that a researcher must be given reasonable time to publish the final report (The committee reference to point 5 in *Vejledende Retningslinjer for Forskningsetik i Samfundsvidenskaberne* from the Danish Social Science Research Council ought really to be re-considered in this light). Rector then demanded of me to be alerted if the promised final publication was not accepted in a peer-reviewed journal. However, this was not enough for to the director of the Institute of Psychology, prof. Jens Mammen, so he overruled the decision of Rector and demanded that not-yet-entirely-finished fragments of the analysis and of the final report be published on my

private homepage at a time when I was under heavy stress while producing a 600+ pages book with a narrow deadline.

- I agree with the committee that the study design for the ASDS is not completely described in the 2005 publication. The reason for this is simple - and unavoidable. The Journal *Personality and Individual Differences* (PAID) has a maximum of 5.000 words for a research article. I actually submitted the article with more details but got it back with instruction to cut it down to the maximum. This left me with a dilemma of providing a full account of all the methodological details but with no space for results and discussion or, alternatively, to present a sparse methodological account, the results and discussion, and leave the option open that methodologically interested researcher could write to the address given in the article for the full methodological account – actually a rather common procedure. I chose the latter option knowing that this limitation applies to all researchers publishing in PAID. I admit, however, that a wiser choice may in this case have been to reduce the other parts of the article in favour of a more detailed methodology section.
- Obviously, the above mentioned space restriction also applies when it comes to a full discussion of problems with repeated measurement and missing data. However, these problems are of primary relevance for the 325 observations sample, and are not at all relevant for evaluation of the 2005 adult sex difference study - for which the data according to the committee are well-defined. The 2003 chapter was primarily intended as a discussion of the outcome of various studies of sex differences. It was never intended to constitute a full report on the final analysis. The latter was indicated in the reference list. It is also worth considering that the references in the 2003 chapter made to the 2001 and 2002 conference presentations were never meant to refer to, or to replace, the data documentation. They were simply mentioned out of respect for the fact that the topic had been brought up previously at two professional meetings and that the reader was entitled to know this. I still have not seen any rules for good scientific research making this kind of referencing questionable. By the way, Pia Ankersen was told already back in January 2002 that the 2001 presentation consisted of notes, tables and figures. She nevertheless talks about "Secret papers" years later.

1.1.3 Procedures used in the analysis of data

- This error is a result of an ambiguous formula in Jensen's (1998) "*The g- factor*" book. I misread it, but it was inconsequential for the adult sex difference.
- In testing the p-value using the standard formula, I was in good faith. I was guided by the following three insights reported in the Jensen (1998) book (p. 538): 1) "The best method for determining the sex difference in psychometric *g* is to represent the sex difference on each of the subtests of a battery in terms of a point-biserial correlation and include these correlations with the full matrix of subtest inter-correlations for factor analysis". In this connection Jensen mentions that 2) "The *g* loading of the sex difference is equivalent to the point-biserial correlation of sex with the test battery's *g* factor …" (p. 540), and also (on p. 542) that "The point-biserial correlation (r_{pbs}) is simply a Pearson product-moment correlation that expresses the relationship between a metric variable (e.g. test scores) and a dichotomous variable (in this case sex, quantitized as male = 1 , female = 0). In a personal communication at the 2001 conference Jensen confirmed that I could use the standard formula for significance testing of the Pearson correlation. I have not been able to get a response from Jensen at the time of writing in order to see whether I have misunderstood him on this point.

- Statsoft provided the software (STATISTICA, version 6) for the hierarchical factor analysis. First, principal components are extracted by a variance maximizing (varimax) rotation of the original variable space. Eigenvalues are calculated by a least squares procedure. The hierarchical factor analytic approach uses a strategy inspired by Thompson (1951), Schmid and Leiman (1957) and Wherry (1959, 1975, 1984; precise references on demand). The clusters of items are identified and axes are rotated through those clusters … "… next the correlations between those (oblique) factors are computed, and the correlation matrix of oblique factors is further factor-analyzed to yield a set of orthogonal factors that divide the variability in the items into that due to shared or common variance (secondary factors) and unique variance due to the clusters of similar variables (items) in the analysis (primary factors)." (Statsoft Manual, Vol. III, p. 3195, 1994.).

After considerable experimentation it was found that the hierarchical factor solution provided the least contamination of *g* from primary factors, which is the reason why I used this approach. It is true, in absolute terms that there will always be an identification problem. However, it is not correct when the committee states that I am apparently unaware of this problem. Thus on page 501 in the 2005 article I mention that the hierarchical factor solution used leads to a *g* factor with "… little dimensional contamination." Moreover, it is possible to demonstrate in practice that the hierarchical approach comes up with a sensible - if not ideal solution. In appendix 2 statistician Bo Sommerlund (in a personal communication to me January 25th. 2002) deliberately biased constructed data for two groups: Group 1 has superior M (e.g. mathematics) and inferior S (e.g. "sprog" or language) scores, and group 2 vice versa and further with lower average. The purpose of this exercise was to see how the hierarchical factor analysis would handle this situation. After the first analysis, Sommerlund performed another hierarchical analysis, this time with 10 M-items and 5 S-items and the ordinary mean, and then a third hierarchical analysis with 5 M-items and 10 S-items and the ordinary mean. He found that the hierarchical factor solution shows impressive robustness of the derived *g* factor against contamination from primary factors.

Way back in 1994 Jensen and Weng raised the question of how invariant *g* is across various methods of factor analysis. They used six different factor analytic methods on four simulated data matrices where the factors were exactly known. They also used nine different factor analytic methods on a real correlation matrix with twenty-four tests taken by 145 grade 7 and 8 students. The average congruence coefficients between the true *g* factor and the *g* factors derived from the various methods amounted to +.998 (range +.997 - +.999). This applied even if some of the artificial matrices were deliberately designed to "trick" deviating estimates. For the real data, the forty-five congruence coefficients between the ten *g* vectors ranged from +.991 to 1,000 (average +.995). They concluded that all the different methods of factor analysis estimated the true *g* so closely that **there was hardly any basis for choosing between them** (my emphasis). Moreover, Ree and Earles (1991) factor analyzed data for 9,173 recruits taking the Armed Services Vocational Aptitude Battery (ASVAB) with 14 different methods and derived 14 different *g* factor scores for each recruit. The many different methods resulted in very little variation among the obtained *g* factors (average correlation +.984). Jensen (1998, p. 83) was able to conclude that "… whatever variation exists among the myriad estimates of *g* that have been reported since the beginning of factor analysis, **exceedingly little of it can be attributed to differences in the methods of factor analysis employed**." (my emphasis).

It was on this empirical basis I decided that providing intimate details about the specific factor analytic approach I used would waste valuable space that was better devoted to more pertinent questions. The reviewers agreed with my decision. I accordingly find that the

committee's insistence on a fuller report of all factor analytic details fails to take into account the empirical evidence in the field.

- 1.1.3/4: It is true that one cannot tell for sure whether an observed difference is due to the secondary or to the primary factors. However, it has been demonstrated empirically that the hierarchical approach used minimizes the contamination of the *g* factor by the primary factors, and **that was all I claimed!** Moreover, many specialists in the field accept this view.

1.2.1 Due diligence

1. Section: the attrition problem. It is incorrect when the committee states that the data acquisition "should have been concluded a decade ago". Obviously, the childhood data collection is over long ago, but not only will missing data still be collected for adult subjects, but further data will be added for years to come. Moreover, there are still raw data for children that are not registered electronically and analysed, and this update procedure will continue despite the committee's claim to the opposite.
2. Section: The 2001 conference presentation was based on 62 persons with complete data for 52 (see 1.1.1 bullit 1).
3. Section: The two errors, one of which was due to an error in using a formula (which I found myself) and a faulty reading of an ambiguous formula, are regrettable, but are of no importance for the conclusion.
4. Section: With respect to the "unusual" size of the dropout problem, see appendix 1. As already said, the actual average dropout rate is about 50%, which is quite common in longitudinal studies. It is also worth noting that most studies of adult sex differences use college students. Such studies will be biased in favour of males due, in part, to their flatter *g* distribution and related overrepresentation at the high g end of the scale as compared to females. In order to minimize this problem, the present sample used elementary school children and young adults as point of departure. This means that they better represent the total population than samples being subjected to increasing harsh selection during high school and university. Jensen (1998, p. 83) thus notes that "The samples most representative of the population are … the studies of elementary schoolchildren randomly sampled from urban, suburban, and rural schools. In other words, the starting point for the present study was less biased than in most other studies.
5. Section: I did discuss the adequacy of some of the common various factor analytic approaches and decided to use hierarchical factor analysis because this method seems to provide a less confounded *g* by group factors than the alternatives.
6. Section: I did make available on request the basis for my 2001 conference presentation (see 1.1.2). It was the press that requested (or rather hunted me down) and afterwards made "extended media statements", some of which I vigorously moderated in public! My response to the press was also dictated by the explicit wishes of the University to generously report ongoing University research. By the way, a sex difference of about the size I found had already been reported repeatedly and long ago in the scientific literature with associated theory. For example, a new study: "Males have greater *g*: Sex differences in general mental ability from 100,000 17- to 18year-olds on the Scholastic Assessment Test" is just presented online in *Intelligence* by Jackson and Rushton. This means that I could actually have responded to the press in exactly the same way exclusively drawing from the research of others, to which I just added my bit of confirmative evidence! As already said (but apparently not noted) Pia Ankersen had all the necessary evidence already back in August 2002 long before she reported to the "Committee for Good Scientific Practice" that I refused

to provide this evidence. In four separate emails I asked for her comments on the material I forwarded to her. Again, she never answered my emails but almost a year later reported to the "Committee for Good Scientific Practice" that I had flatly refused to provide her with material and to cooperate.

1.2.2 Mistakes

1. Section: Again, the use of the standard hypothesis test for testing the point-biserial correlation coefficient was made after an advice by professor Arthur Jensen, who sees the point-biserial correlation as simply a Pearson product-moment correlation" (Jensen, p. 542, note 9).

3. section: The committee mentions two "… clear mathematical facts that Nyborg does not seem to be aware of." First there is an inherent unidentifiability …" … Second, Nyborg claims that his version … avoids the problem …". Neither is correct. Nowhere do I claim that my version of the *g*-factor method avoids the inherent unidentifiability problem nor that it avoids the problem. What I claim is rather that this approach minimises the contamination problem relative to other types of factor analyses (see 1.1.3 bullit 5).

The identifiability problem is a controversial and highly technical question of relevance for a broad professional audience. I therefore find it more appropriate that the committee published their analysis of the problem in the "Personality and Individual Differences" journal, where it would receive the qualified treatment it deserves. I find it entirely unacceptable that the committee uses its highly specialised insights in complex mathematics and statistics against a scientist in a closed university hearing where his position as professor of psychology is explicitly put at stake!

5. Public access

On page 11 the committee mentions the SSRC rules that "… conclusions should not be made public … before the investigation is concluded and made available. What then if a journalist demands to learn about a study that has received public money? Should conference organizers stop urging the scientists to go to the newsroom and let them be interviewed about the most recent results that have not been published before (as required by many international conferences)? Should the university stop calling for researchers going public with results that interests the public even if the conclusions are preliminary as is often the case (I believe that most congress papers are modified before finally being published, some radically in the peer review process). I see no clear rules here safely guiding a researcher that is under constant and increasing pressure to make public ongoing research that has already been made open for inspection at professional conferences! What I am saying is that much common practice does not conform entirely to the SSRC rule.

6. Reconstructing Nyborg's results

I note that the committee "…generally get[s] similar results as Nyborg, using a broad set of related methods." and further that the data sets used for publication are "… reasonably well defined."

All things considered, it it rather difficult to see why the above mentioned problems can lead to such serious consequences.

References

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

Jensen, A.R., & Weng, L-J. (1994) What is a good *g*? *Intelligence, 18*, 231-258.

Ree, M.J., & Earles, J.A. (1991) The stability of convergent estimates of *g*. *Intelligence, 15*, 271-278.