

L'ECHO DES MESSACHES – Novembre 1979 – N°9

En marge de l'affaire Burt :

CYRIL BURT AND THE CARELESS STAR WORSHIPPERS par Louis Guttman

On connaît l'"affaire Burt": l'illustre psychométricien britannique Cyril Burt, accusé (à titre posthume, il était mort depuis cinq ans) à la suite d'un retentissant article du *Sunday Times* (1), du crime suprême que puisse commettre un chercheur scientifique : celui d'avoir fabriquer ses données. Le débat qui s'en est ensuivi, d'autant plus âpre qu'il est venu se greffer sur la controverse à la mode sur l'hérédité du QI etc., a connu des rebondissements divers : véritable enquête de détective (« les deux collaboratrices de Burt, qui étaient censées recueillir les données et qui cosignaient ses articles, ont-elles seulement existé ? ») et articles de revues scientifiques, se proposant d'expertiser, « en toute objectivité », les tableaux de données incriminés. C'est ainsi qu'a paru en 1978, dans la prestigieuse revue *Science*, un volumineux article, dans lequel sont scrutées les anomalies (ou plutôt, en l'occurrence, les trop belles propriétés de « normalité ») de tableaux de contingence présentés par Burt dans son article : « *Intelligence and social mobility* ». A grand renfort de tests de signification (2), l'auteur de l'article de *Science* (D. Dorfman) conclut qu'il est hors de doute (« beyond reasonable doubt ») que Burt a « fabriqué les données de ces tableaux ».

C'est à ce point que se situe le très intéressant texte de Guttman *Cyril Burt and the careless star worshippers* (3) dont on lira les extraits ci-dessous. A propos de l'article paru dans *Science* et par delà le cas « dramatique » de l'« affaire Burt », le célèbre sociologue dénonce l'usage illégitime des tests de signification, trop souvent invoqués dans les recherches à prétention scientifique, comme critère d'administration de la preuve. La critique de Guttman, dont la portée est tout à fait générale, rejoint celles qui depuis des années sont émises par les chercheurs les plus éminents et les plus avertis (4), et qui ont été naguère évoquées dans un article de l'*Echo* (5).

H. Rouanet

(1) Voir *Le Monde* daté du 12.01.1977

(2) Significatif à $P=.05$: une étoile (*) ; à $P=.01$: deux étoiles (**), etc.

(3) Ce texte distribué par le Professeur Guttman lors d'une conférence donnée à Paris au printemps dernier, est disponible sur demande.

(4) Voir par exemple le recueil d'articles Morrisson & Henkel : *The significance test controversy*, London, Butterworths, 1970.

(5) « Le Khi deux, le Khi deux, mais qui d'eux ? », n°5, mars 1976.

CYRIL BURT AND THE CARELESS STAR WORSHIPPERS

Louis Guttman

It is of interest to find Cyril Burt under posthumous attack for careless presentation of certain data. There may be grounds for such criticism, but Burt's posthumous critics are perhaps just as careless – or even more careless and more unscientific – than the late psychologist in discussing these particular data.

A bibliography of attacks on Burt's presentation of data with alleged "new findings" is to be found in a recent paper by Dorfman¹, in the prestigious journal *Science*; for the "new findings" it uses one of the supposedly most sophisticated tools of proof known to science: tests of statistical significance. In particular, probabilities calculated from the distribution of the chi-square statistic are used to show "beyond reasonable doubt" that Burt tampered with his data before publishing them.

I do not doubt at all that Burt adjusted his data to arrive at the four tables reproduced in the Dorfman article. At the same time, I do not doubt at all the several probabilities Dorfman has so laboriously calculated and published as "new findings" are not only adjusted, but are wrong.

Dorfman is in good company in having carelessly committed two major errors – these are endemic among psychologists and other behavioral scientists. First, he calculates as if his article is concerned with only one test of significance, whereas in fact it deals with *several simultaneous* hypotheses. The single-hypothesis probabilities he presumes to hold for his calculations are incorrect. Second, even if these probabilities were correct, they would be so for the wrong problem, for Dorfman has committed the standard error of reversing the role of null and alternative hypotheses.

Along with many other researchers, Dorfman has fudged the whole concept of tests of significance. Jerzy Neyman, who together with Egon Pearson founded the modern school of inference for hypothesis testing, is himself still looking for a solution to the problem of simultaneous hypotheses. For a review of the inferential problems of simultaneous hypotheses (multiple comparisons) see, for example, Nemenyi². The problem of proper choice of null hypothesis is an additional complication. For a discussion of these and other matters belonging to "what is not what in statistics" see Guttman³. For an example of alternative discussions of fallacies of the null hypothesis approach, see Rozeboom⁴.

Anyone who knows that Burt was a mental tester *par excellence*, and who sees the smoothness of Burt's tables, should conclude immediately that Burt adjusted his data to approximate some neat (bivariate normal) distribution. Burt may have discussed only the adjusted distribution for pedagogical reasons. It appears he was careless in not always pointing out that he was representing the original data only schematically. All this has nothing to do with the truth or falsity of his hypothesis.

It is about time that *Science* and other journals stopped accepting articles that purport to be "scientific" and "hard-headed" when such articles merely embellish themselves with so-called tests of significance. Some of my colleagues tell me that, for publication purposes, they feel forced to pepper their manuscripts with stars and double-stars (for the supposed .05 and .01 levels of significance) or else editors will reject their mss. My colleagues know that such star worship is but a miserable masquerade. They try to stifle their feelings of scientific conscience by the excuse that "everyone is doing it".

Publishing pseudo-statistical inference doesn't speak well for scientific competence, or even for scientific integrity (of those who feel "forced" to conform). Many journals have long backlogs of papers accepted for publication. These backlogs might be cut down drastically if star worshipping papers were given low or no priority. Considerable amounts of scarce research funds could also be saved, by rejecting research projects based essentially on star worship.

¹ D.D. DORFMAN, *Science*, 209, 1177, (1978).

² P. DEMENYI, Multiple comparisons, *International Encyclopedia of the Social Sciences*, 9, 337, (1968).

³ L. GUTTMAN, *The statistician*, 26, 81, (1976).

⁴ W. W. ROZEBOOM, *Psychol. Bull.*, 57, 416, (1960).